Relative Valuation with Machine Learning

Paul Geertsema* and Helen Lu†

First version: 29th May 2019

This version: 28th August 2022

Abstract

We use machine learning for relative valuation and peer firm selection. In out-of-sample tests our machine learning models substantially outperform traditional models in valuation accuracy. This outperformance persists over time and holds across different types of firms. The valuations produced by machine learning models behave like fundamental values. Over-valued stocks decrease in price and under-valued stocks increase in price in the following month. Determinants of valuation multiples identified by machine learning models are consistent with theoretical predictions derived from a discounted cash flow approach. Profitability ratios, growth measures and efficiency ratios are the most important value drivers throughout our sample period. We derive a novel method to express valuation multiples predicted by our machine learning models as weighted averages of peer firm multiples. These weights are a measure of peer-firm comparability and can be used for selecting peer-groups.

Keywords: Relative Valuation; Peer Firms; Fundamental Analysis; Machine Learning; GBM; Gradient Boosting Machine

JEL codes: G12; G14; M41; C5

*E-mail: p.geertsema@auckland.ac.nz. Tel: +64 9 923 9431. Department of Accounting and Finance, University of Auckland Business School, Owen G Glenn Building, 12 Grafton Road, Auckland, New Zealand. Postal Address: The University of Auckland Business School, Private Bag 92019, Auckland 1142, New Zealand.

†Send correspondence to helen.lu@auckland.ac.nz. Tel: +64 9 923 3326. Department of Accounting and Finance, University of Auckland Business School, Owen G Glenn Building, 12 Grafton Road, Auckland, New Zealand. Postal Address: The University of Auckland Business School, Private Bag 92019, Auckland 1142, New Zealand.

Electronic copy available at: https://ssrn.com/abstract=3740270

We are grateful for the insightful comments from the editor (Haresh Sapra) and two anonymous referees. We thank participants at the Department of Accounting and Finance seminar series at the University of Auckland (2019 and 2020), the Quantitative Methods in Finance conference (Sydney, 2019), the 32nd Australasian Finance and Banking Conference (Sydney, 2019), the 24th and 25th Annual New Zealand Finance Colloquium (NZFC, Auckland, 2020 and Tauranga, 2021) and an event organised by the Institute of Finance Professionals New Zealand Inc (INFINZ, 2020) for their valuable comments. An earlier version of this paper was awarded the INFINZ Best Paper for Investment award at the 24th New Zealand Finance Colloquium. We thank Pedro Barroso, Henk Berkman, Steven Cahan, Griffin Geng, Maryam Hasannasab, David Hay, Stephen Kean, Paul Healy, Michael Keefe, Robert Knechel, Nick Nguyen, Peter Philips, Paul Rouse, Derek Snow, Charl de Villiers, Jilnaught Wong and Norman Wong for insights and suggestions. Any remaining errors are our own.

Electronic copy available at: https://ssrn.com/abstract=3740270

1 Introduction

Relative valuation is widely used by finance practitioners. For example, relative valuation is used in investment bankers' fairness opinions (Bartell & Janssen, 2017), in valuing initial public offerings (IPOs) (Kim & Ritter, 1999) and by sell-side research analysts (De Franco, Hope & Larocque, 2015). Relative valuation asks a very simple question: is a stock overvalued or undervalued relative to its peers in the *current* market? The choice of peers is often highly subjective and loosely based on "industry". Both empirical and anecdotal evidence suggests that practitioners strategically select peers to achieve desired valuation results (e.g. De Franco, Hope and Larocque, 2015 and Eaton, Guo, Liu et al., 2021).

Academics have proposed a number of methods to add rigour to peer selection and relative valuation. Hoberg and Phillips (2010, 2016) introduce a text-based industry classification method. These industry classifications are based on pairwise firm similarity scores derived from textual analysis of firms' business descriptions filed with the Securities and Exchange Commission (SEC). Lee, Ma and Wang (2015) use investor's sequential search activities on the SEC's EDGAR website to infer how closely firms are linked with each other. Relative valuation can potentially benefit from these new peer firm selection algorithms. However, the peer selection step can be bypassed if we estimate valuation multiples directly using firms' fundamentals. In fact, academics have long advocated estimating valuation multiples using firm characteristics through ordinary-least-squares (OLS) regressions (e.g. Bhojraj and Lee, 2002; Cragg and Malkiel, 1968; Damodaran, 2007; Whitbeck and Kisor Jr, 1963).

The relationship between fundamentals and valuation is well motivated by finance theory. However, prior studies often use a limited number of variables and assume a linear relationship between those variables and a valuation multiple. In practice, investors rely on a large number of variables which interact in complex and non-linear ways to determine a firm's value. Recent advances in machine learning can assist in disentangling these often-nuanced relationships, thereby providing more accurate relative valuations and more objective peer firm selections. In this paper, we propose a machine learning approach for relative valuation and peer selection. This approach does not require a pre-defined theoretical model. Instead, it identifies fundamentals that are important for relative valuation to predict valuation multiples out-of-sample. The specific machine learning approach used in this paper is gradient boosting machines (GBMs), which is a tree-based machine learning model.[1] GBMs can make accurate predictions without over-fitting the data when the underlying relationship is complex and not specified in advance (Mullainathan & Spiess, 2017).

[1] We discuss tree-based models in detail in section 3.1

Electronic copy available at: https://ssrn.com/abstract=3740270

We task the GBMs with predicting the natural logarithms of market-to-book (*m2b*), enterprise-value-to-assets (*v2a*) and enterprise-value-to-sales (*v2s*) multiples (hereafter, *targets* or *target variables*) because the denominators used in these multiples are available and non-negative for most firms in the Compustat universe. Our results show that GBM models generate more accurate out-of-sample predictions than traditional models (Bartram and Grinblatt, 2018; Bhojraj and Lee, 2002; Hoberg and Phillips, 2010, 2016; Liu, Nissim and Thomas, 2002; Rhodes-Kropf, Robinson and Viswanathan, 2005). Equity valuations from GBMs have substantially lower median absolute valuation errors, smaller root mean squared errors and higher out-of-sample R-squared's when benchmarked against traditional models. For example, among the traditional models, peer-group mean models based on Hoberg and Phillips (2010, 2016) (*HP* models) generate the lowest median absolute valuation errors (ranging between 0.40 and 0.59), but these are still much higher than those from machine learning models (ranging between 0.29 and 0.32). Based on root mean squared errors and out-of-sample R-squared's, the best performing traditional models are the regression models proposed by Bhojraj and Lee (2002) (*BL* models). Root mean squared errors from *BL* models range from 11.4 to 12.4 billion US dollars, or about 4 billion US dollars higher than those from machine learning models. Out-of-sample R-squared's from *BL* models range between 0.38 and 0.48, while machine learning models improve this measure to between 0.76 and 0.78.

The outperformance of GBMs over traditional models is ubiquitous. GBMs generate substantially lower median absolute valuation errors than *any* of the traditional models *in every month* throughout the sample period (1980-2019). GBMs also generate more accurate valuations than traditional models across different size, market-to-book and ROE quintiles.

To validate that GBM valuations are a good proxy for fundamental value, we test if future equity value converges to the valuation result. If a valuation prediction closely resembles fundamental value, the stocks that the machine deems to be over-valued should decrease in price while those deemed under-valued should increase in price. We find that machine valuation errors strongly predict future returns. This suggests that GBM valuations are well-aligned with fundamental value.

In addition to generating accurate out-of-sample predictions, results from GBMs can also be used to select peer firms with similar fundamentals. A GBM model consists of sequentially constructed trees. Firms are allocated to different "leaves" in a tree based on decision rules using firm fundamentals. Firms that are often allocated to the same leaf can be seen as close peers with similar fundamentals. Conversely, firms that are rarely allocated to the same leaf

Electronic copy available at: https://ssrn.com/abstract=3740270

have dissimilar fundamentals. In fact, we show that a GBM predicted valuation multiple can be expressed as a weighted average of valuation multiples of all firms in the training data. These weights can be interpreted as a measure of comparability. Firms with high weights are closely comparable companies while those with low weights are dissimilar companies. Peer weights can be used to scrutinise the choice of companies used in relative valuations. Our decomposition of GBM predictions into equivalent peer weights is a novel contribution to explainable AI.

Since machine learning algorithms take a data-driven, atheoretical approach to construct models, it is necessary to validate whether the machine learning approach is congruent with finance theory. To facilitate the analysis, we start with the Gordon constant growth model (Brennan, 1971; Gordon, 1962) to express each valuation multiple as a function of basic accounting ratios. We show that GBMs rely on ratios that directly relate accounting amounts (book, assets or sales) to profits; that is, the models "learned" from the data appear to conform to theory. For example, ROE-type variables are more dominant for equity-value target multiples while ROA-type variables are more dominant for enterprise-value-to-asset target multiples. Contrary to the common practice of choosing firms from the same industry as peers, industry classification is not the most important category in the machine valuation process. We find that variables in the Profitability, Growth, Financial soundness and Efficiency categories are all more important than industry in predicting valuation multiples.

Several recent papers have applied machine learning techniques to topics in accounting and finance. These include stock return predictability (Gu, Kelly & Xiu, 2020), analysis of the textual structure of news (Bybee, Kelly, Manela et al., 2020), bond yield predictability (Bianchi, Büchner & Tamoni, 2020), credit risk forecasting (Khandani, Kim & Lo, 2010), dimensionality of risk factors (Geertsema & Lu, 2020), fraud detection (Bao, Ke, Li et al., 2020) and director selection (Erel, Stern, Tan et al., 2021). Our application of machine learning to relative valuation is novel. We contribute to the literature on relative valuation (e.g., Bartram and Grinblatt, 2018; Bhojraj and Lee, 2002; Golubov and Konstantinidi, 2019; Liu, Nissim and Thomas, 2002), the literature on peer firm selection (e.g., Bhojraj and Lee, 2002; Hoberg and Phillips, 2010, 2016; Lee, Ma and Wang, 2015) and the literature on the value relevance of accounting information (e.g., Barth, Li and McClure, 2021; Brown, Lo and Lys, 1999; Collins, Maydew and Weiss, 1997; Core, Guay and Van Buskirk, 2003; Francis and Schipper, 1999; Lev and Zarowin, 1999).

The following section reviews the literature. Section 3 describes our GBM models and the sample selection process, section 4 reports model performance, section 5 identifies influential valuation drivers and section 6 concludes.

Electronic copy available at: https://ssrn.com/abstract=3740270

2 Related literature

Traditional relative valuation models can be broadly categorised into industry mean approaches, fundamental analysis or a combination of these two (see Damodaran (2007) for a review). We select several high-profile traditional models as benchmarks for our machine learning models. Section IA.1 in the Internet Appendix summarises the variables used in each of the traditional models.

Liu, Nissim and Thomas (2002) take a pure industry approach. They find that industry harmonic means provide more accurate valuation multiple estimates than industry medians and means. We use this harmonic mean approach to estimate all three valuation multiples and refer to these approaches as *IND_m2b*, *IND_v2a* and *IND_v2s* (*IND* stands for industry).

Hoberg and Phillips (2010, 2016) analyse the similarity of business descriptions in annual reports filed with the SEC to assign pair-wise similarity scores between firms in each year. We use their similarity score matrix[2] as a starting point to identify peers of each firm. For a given firm, we select the 60 firms with the highest similarity scores in that year as its peer firms.[3] The estimated valuation multiple is the average of current valuation multiples of peer firms based on prior year annual reports. We refer to these three approaches as *HP_m2b*, *HP_v2a* and *HP_v2s*.

By contrast, Bartram and Grinblatt (2018) uses a pure fundamental approach to estimate the relative fundamental value of firms at a point in time. They estimate equity value by regressing the actual market value of all stocks on 28 accounting items in each month. We refer to the Bartram and Grinblatt (2018) model as the *BG* model.

Many papers use both industry classification and accounting information to estimate the fundamental value of stocks or valuation multiples. We choose models from two influential papers, Rhodes-Kropf, Robinson and Viswanathan (2005) and Bhojraj and Lee (2002) as additional benchmarks.

Rhodes-Kropf, Robinson and Viswanathan (2005) put forward a model based on residual income theory. It predicts the log of equity value using industry-month regressions on fundamental variables. While they investigate deviations from fundamental value both in the cross-section and over time, we only use their cross-sectional regressions to estimate relative

[2] https://hobergphillips.tuck.dartmouth.edu/industryclass.htm

[3] We select 60 firms to be consistent with other models in this study because the number of firms in an FF49 industry averages around 60 in our sample. A minimum of 5 non-missing peer firms is required for a valid Hoberg and Philips prediction.

Electronic copy available at: https://ssrn.com/abstract=3740270

fundamental value, so that all models studied in this paper are comparable. We refer to this cross-sectional valuation model as *RRV*.

Our paper is most closely related to Bhojraj and Lee (2002) (*BL*) in that we both emphasise the importance of guiding the choice of peer firms with fundamentals, instead of focusing solely on industry classification. We refer to these three models based on Bhojraj and Lee (2002) as *BL_m2b*, *BL_v2a* and *BL_v2s*.

Bhojraj and Lee (2002) motivate their selection of model inputs with finance theory. We instead use a data-driven approach and demonstrate that GBMs can identify important valuation drivers congruent with theory from a large number of variables. The important valuation drivers identified by GBMs encompass a number of explanatory variables in *BL* models including R&D, growth, leverage and accounting returns.

We also contribute to the value relevance literature. This literature initially focused on the ability of earnings and book equity to explain equity value (Brown, Lo & Lys, 1999; Collins, Maydew & Weiss, 1997; Francis & Schipper, 1999). Barth, Li and McClure (2021) use the Classification and Regression Tree (CART) approach introduced by Breiman, Friedman, Olshen et al. (1984) to study the evolution of relations between accounting variables and equity value. Both Barth, Li and McClure (2021) and our study use tree-based approaches.[4] Their primary interest is the evolution of value relevance of accounting variables whereas our study investigates the ability of machine-learning methods to perform firm valuation and peer selection.

3 Research design

3.1 Gradient boosting machines

Our study uses gradient boosting machines (GBMs) to predict valuation multiples out-of-sample.[5] A GBM is a decision-tree based machine learning method that generates accurate

[4]Both CART and GBMs are tree-based approaches. GBMs are the current state-of-art machine learning approach in the context of tabular data (Grinsztajn, Oyallon & Varoquaux, 2022; Ke, Meng, Finley et al., 2017). GBMs adds trees in an iterative process that improves accuracy while regulating over-fitting. By contrast, CART is simply a single tree predictor. In our setting GBMs generate substantially more accurate predictions than CART (see section 4.1).

[5]GBMs have been used in finance research on return prediction, director turnover prediction and stock valuation (e.g., Cao, Jiang, Wang et al., 2021; Erel, Stern, Tan et al., 2021; Gu, Kelly and Xiu, 2020). Section A.2 in the Appendix provides an in-depth explanation of GBMs.

Electronic copy available at: https://ssrn.com/abstract=3740270

and robust out-of-sample predictions without over-fitting the data.[6] This type of model learns the optimal decision rules from the training (in-sample) data to construct several decision trees. It iteratively builds many trees to improve prediction accuracy. A GBM model uses these decision rules to allocate each test (out-of-sample) firm to a leaf on each tree to make an out-of-sample prediction. All results presented in this paper are from predicted valuation multiples in *out-of-sample* tests. Section A.2 in the Appendix presents a simplified example of how GBMs generate out-of-sample predictions.

We follow the standard machine learning practice of splitting our dataset into three non-overlapping datasets in each month: a *training* dataset to train the GBM, a *validation* dataset to select the optimal number of trees and a *test* dataset to evaluate out-of-sample performance.[7] In our study, we choose a cross-sectional split of 60:20:20 for the training, validation and test datasets by month. In each month, observations are randomly allocated to one of five "blocks". Each block takes a turn to be the test sample to obtain their out-of-sample predictions. Using these rotations, we generate out-of-sample predictions for all observations in a month. Hence five models are trained every month for each of the 480 monthly cross-sections in our sample and this is done for each of the three target multiples. As a result, we train a total of 7,200 separate machine learning models (5 test data blocks/month $\times$ 480 months $\times$ 3 target multiples = 7,200 models). To make the results comparable across models, traditional models also use a similar cross-sectional split of data in each month to make out-of-sample predictions. The process is illustrated graphically in Figure IA.1 in the Internet Appendix. We also consider an alternative sample split of 80:20 for the training and test datasets by month and our main findings are robust to this alternative sample split (see section IA.6.1 in the Internet Appendix).

To perform relative valuation using GBMs we use three valuation multiples: market-to-book (*m2b*), enterprise-value-to-assets (*v2a*) and enterprise-value-to-sales (*v2s*). These three machine learning models are denoted by *ML_lnm2b*, *ML_lnv2a* and *ML_lnv2s*. To mitigate the impact of outliers in target variables, GBMs in our study target log multiples (*lnm2b*,

[6] In online machine learning competitions using true out-of-sample data such as those hosted by Kaggle (www.kaggle.com), GBMs dominate solutions involving structured data (Chen & Guestrin, 2016). Two high-profile implementations in this context are *XGBoost* (Chen & Guestrin, 2016) and *LightGBM* from Microsoft Research (Ke, Meng, Finley et al., 2017). Both XGBoost and LightGBM implement numerous algorithmic optimisations aimed at increasing accuracy and speed, which sets them apart from the earlier generation of gradient boosting machines based on the seminal work of Friedman (2001, 2002). We use LightGBM version 3.2.1 from Microsoft Research (Ke, Meng, Finley et al., 2017).

[7] The number of trees (`num_boost_round`) are chosen via early stopping on the validation dataset. For all other hyper-parameters we use the default settings of *LightGBM*; these default settings include a learning rate (`learning_rate`) of 0.1, a maximum of 31 leaves per tree (`num_leaves`) and a minimum of 20 observations per leaf (`min_data_in_leaf`).

Electronic copy available at: https://ssrn.com/abstract=3740270

lnv2a and *lnv2s*).[8]

To conduct a relative *fundamental* valuation, we limit input variables to those based on (non-market) fundamental information only (with the exception of the CAPM beta to proxy for systematic risk). We use all accounting ratios in the WRDS ratio suite (numbering 52 and indicated with a "_wr" suffix), sorting variables for accounting anomalies from the prior literature (numbering 39 and indicated with an "_an" suffix), four accounting items to proxy for size (total assets, book equity, sales and book value of total debt), the CAPM beta and the Fama-French 49 industries classification. Section A.1 in the Appendix describes the construction of these input variables and section IA.2 in the Internet Appendix reports summary statistics. Some accounting ratios convey similar information but are constructed differently. Thus, to facilitate interpretation, we group these 97 input variables into 11 *categories* (see section IA.2 in the Internet Appendix for categories) closely following the categories used in the WRDS ratio suite.

3.2 Sample selection

We source our data from the CRSP monthly and Compustat quarterly databases. We include all US common equities (*shrcd* 10 or 11) listed on NYSE, NASDAQ and AMEX (*exchcd* 1, 2 or 3) that have non-missing market equity values and with total assets, sales and book equity each above the 10th percentile in each month.[9] Our sample starts in January 1980 and ends in December 2019. The final sample used by machine learning models consists of 1,811,785 firm-month observations, which correspond to 16,201 unique firms (in terms of *gvkey*). Table 1 summarises the sample selection steps.

[Insert Table 1 about here]

Using GBMs allows us to impose only minimal filtering during sample selection. For example, LightGBM automatically imputes missing values in input variables; hence, observations

[8] GBM predictions are sensitive to outliers in the *target variable* because the prediction is the average of target values of training observations allocated to the same leave. In contrast, GBMs are robust to outliers in *input variables* because a decision rule splits data based on a threshold which is robust to outliers.

[9] We use accounting variables instead of the equity value to exclude tiny firms in order to avoid information leakage from the target variable (market-based valuation multiples) to the input dataset. It is important to note that this filtering is only applied to the data used to train the machine learning models. Specifically, when predicting returns using valuation errors (in section 4.3), we use all available returns without any filtering on firm size to avoid look-ahead bias.

Electronic copy available at: https://ssrn.com/abstract=3740270

with missing values for input variables are included in our study. In addition, we do not winsorise any variables because decision rules in GBMs use thresholds and are therefore robust to outliers in input variables. Neither do we exclude firms in the finance industries or regulated industries, as is common in many empirical accounting and finance studies. This is because tree-based models automatically capture non-linear and multi-level interaction effects via applying sequential decision rules. In this way GBM models can overlay industry classification with decision rules based on other fundamental information.

At the end of each month, both machine learning models and traditional models use information available at the time in the training set to predict valuation multiples out-of-sample in the test dataset. *Month* in our datasets always refers to the valuation month. We assume that all accounting information becomes available three months from the balance-sheet date, following Hou, Xue and Zhang (2018).[10] Traditional models and machine learning models use the same datasets, except that traditional models require that an observation has non-missing values for all independent variables to be included in a regression.

4 Model performance

4.1 Performance benchmarking

We use several metrics to evaluate model performance. These metrics include percentage valuation errors of the predicted value of equity (ε), root mean squared errors (RMSE), out-of-sample R-squared's (r^2_{oos}) and the Pearson correlation between actual and predicted equity values ($\rho_{M,\hat{M}}$). The percentage error measures the accuracy in predicting the equity value and is defined as $\varepsilon := \hat{M}/M - 1$ where M is the equity value and $\hat{M}$ is the corresponding model estimate. The equity value (M) is the absolute value of price (*altprc*) times the number of shares outstanding (*shrout*) at the end of each month in CRSP. An alternative measure of prediction accuracy is RMSE, which measures the difference between the actual values of a variable (x) and predicted values of a variable ($\hat{x}$), and is defined as $rmse(\hat{x}) := \sqrt{\frac{1}{N}\sum(x-\hat{x})^2}$. We also report out-of-sample R-squared's (R^2_{oos}). R^2_{oos} measures how well out-of-sample predictions ($\hat{p}$) perform relative to simply using the mean of training sample values ($\bar{p}$) as a prediction, defined as $R^2_{oos}(\hat{p}) := 1 - \frac{\Sigma(p-\hat{p})^2}{\Sigma(p-\bar{p})^2}$.

[10] Hou, Xue and Zhang (2018) assumes that accounting information becomes available for use at the end of three months after the balance sheet date, for use in predicting returns in the fourth month.

Electronic copy available at: https://ssrn.com/abstract=3740270

We report the *dollar* RMSE ($rmse(\hat{M})$) and the *dollar* out-of-sample R-squared ($R^2_{oos}(\hat{M})$), alongside the *model* RMSE ($rmse(\hat{y})$) and the *model* R-squared ($R^2_{oos}(\hat{y})$). A *dollar* metric considers the dollar equity valuation while a *model* metric uses the target variable specific to that particular model. We use *dollar* metrics to compare the performance of models with different targets when *model* metrics are unsuitable (for example, the *ML* models targeting the logs of valuation multiples versus the *BG* model targeting market capitalisation). Section IA.4 in the Internet Appendix provides additional detail regarding the translation of predicted target values to predicted equity values.

Table 2 reports the performance of all the models. Machine learning models (those prefixed with "*ML*") outperform traditional models by a substantial margin.

[Insert Table 2 about here]

First, machine learning models generate median absolute valuation errors (med($|\varepsilon|$)) ranging between 0.29 and 0.32, *all* of which are lower than those produced by *any* of the traditional models. Machine learning models predicting the log of market-to-book multiple (*ML_lnm2b*) generate the lowest med($|\varepsilon|$), followed by *ML_lnv2a* and then *ML_lnv2s*. The lowest med($|\varepsilon|$) from traditional models is 0.40 from the *HP_m2b* model. The *dollar* RMSE ($rmse(\hat{M})$) from machine learning models range between \$7,289 mn and \$7,537 mn, which is approximately \$4,000 mn lower than that of the next best-performing model (*BL_v2a* in Table 2, at \$11,414 mn). Similarly, the *dollar* out-of-sample R-squared's ($R^2_{oos}(\hat{M})$) for ML models range between 0.76 to 0.78, exceeding the best traditional model (*HP_m2b*, 0.51) by about 0.25.

Second, machine learning models achieve consistent accuracy across a range of target valuation multiples. The ranges of valuation errors, RMSE's and out-of-sample R-squared's are very similar across all three machine learning models in Table 2. By contrast, the performance of traditional models varies substantially depending on the valuation multiple targeted – for instance, *IND_m2b* yields a $R^2_{oos}(\hat{M})$ of 0.38 but *IND_v2s* yields a substantially worse $R^2_{oos}(\hat{M})$ of -0.22. This demonstrates the flexibility of the GBM machine learning approach in automatically picking the best internal specification for each target multiple.

Third, machine learning models also generate the least biased equity valuations. Median valuation errors (med(ε)) of the three machine learning models are close to zero. The *BL* models tend to overvalue stocks, as indicated by their large and positive median valuation errors (between 20% and 60%). By contrast, both *IND* models and *HP* models tend to

Electronic copy available at: https://ssrn.com/abstract=3740270

undervalue stocks, generating negative median valuation errors (between −15% and −35% and between −16% and −22%, respectively).

Finally, machine learning models also generate equity valuations that are highly correlated with the actual market values of equity. The correlations ($\rho(M, \hat{M})$) range between 0.89 and 0.90 for the three machine learning models. This is followed by *BL_v2a* (at 0.86) and *BL_v2s* (at 0.83).

In Table 2, GBM models generate substantially more predictions than traditional models because GBMs accommodate missing values in the input data while traditional models do not. Machine learning models continue to outperform traditional models in terms of accuracy when restricted to a sub-sample of test data with valid predications from all models (see Table IA.5 in the Internet Appendix).

We also introduce a few variations to the base machine learning models. These variations include using alternative targets in GBMs, using CART instead of GBMs and applying an alternative sample split to train GBMs. *First*, GBMs targeting log valuation multiples (or, the base ML models) outperform those targeting equity value, log equity value or valuation multiples (section IA.6.1 in the Internet Appendix). This outperformance persists over time and holds across different firm types (Figure IA.2 in the Internet Appendix). *Second,* CART models are generally superior to the non-machine-learning models, but underperform the GBM models (*ML_lnm2b, ML_lnv2a, ML_v2s*) for the same target variable across every performance metric (section IA.6.2 in the Internet Appendix). *Third*, under an alternative data splitting regime for the GBM models the results are marginally weaker than the base ML models (*ML_lnm2b, ML_lnv2a, ML_v2s*) but still significantly stronger than the traditional approaches for each of the target variables (section IA.6.3 in the Internet Appendix).

4.2 Model performance over time and across firms

GBMs produce substantially smaller absolute valuation errors than traditional models in every single month in our sample period. Panels A to C in Figure 1 compare median absolute valuation errors from the machine learning models to those from traditional models targeting the same valuation multiple. Panel D in Figure 1 plots median absolute valuation errors from all machine learning models and those from *RRV* (targeting the log of equity value) and *BG* (targeting equity value). In all graphs, lines depicting median valuation errors from machine learning models always lie below the lines for errors from traditional models.

Electronic copy available at: https://ssrn.com/abstract=3740270

These graphs also show that valuation is more challenging in certain markets for all models. Median absolute valuation errors from all models were elevated during the internet bubble in 2000 and again during the 2008 financial crisis, but less so for the machine learning models.

[Insert Figure 1 about here]

Panel E of Figure 1 compares the performance of the machine learning models amongst themselves. While the model predicting the log of the market-to-book multiple (*ML_lnm2b*) produces more accurate equity valuations on average throughout the sample period, its superior performance has weakened over time as shown by the narrowing gaps between the three lines. This finding is likely to be related to the declining ability of book equity to capture firms' economic assets (Enache & Srivastava, 2018) because the economy has become more knowledge and service centric.[11]

GBMs also generate more accurate valuations than traditional models across different size, market-to-book and ROE quintiles. In Figure 2, we group firms into quintiles in each month based on their market capitalisation, market-to-book ratios and ROEs, before plotting the median absolute valuation error for each group. Lines depicting median absolute errors from machine learning models are almost always well below the lines from other models across all quintiles. However, for certain types of firms, some traditional valuation methods and the GBM models generate similar median absolute valuation errors. For example, for firms in quintiles two and three of market-to-book ratios, median absolute valuation errors from *IND* models are similar to those from GBMs for all three multiples (Panel A.2. in Figure 2 and Panel A.2 in Figures IA.4 and IA.5 in the Internet Appendix). In addition, for firms in the highest market-to-book ratio quintile, the regression approach based on Bhojraj and Lee (2002) and GBMs generate similar median absolute errors for all three multiples.

[Insert Figure 2 about here]

Which firms are harder to value for machine learning models? Figure 3 answers this question.

In the cross-section, small firms, firms with extreme market-to-book ratios and firms with low returns on equity (ROEs) are harder to value than other firms (Panels A.1 to A.3 of

[11] In section IA.7 of the Internet Appendix we show that Old Economy firms account for all of the superior performance of the book-to-market valuation multiple with no discernible difference evident among New Economy firms. Machine learning models also generate more accurate predictions for Old Economy firms compared to New Economy firms (see Panels C and D in Figure IA.3 in the Internet Appendix). This apparent disconnect between accounting information and capital market valuation in the context of New Economy firms is also documented in Donelson, Jennings and McInnis (2011) and Srivastava (2014).

Electronic copy available at: https://ssrn.com/abstract=3740270

Figure 3). Small firms and firms with low ROEs have large variations in their valuation multiples and are intrinsically harder to value. Firms in the extreme quintiles sorted by market-to-book ratios are most likely undervalued or overvalued; the large median absolute valuation errors in these two quintiles reflects this. Across the Fama-French 12 industries, firms in the energy, healthcare and telecom industries are more difficult to value than those in other industries, indicated by the peaks of median absolute valuation errors for these three industries (Panel A.4 of Figure 3). By contrast, financial firms (*Money*) and utilities firms (*Utils*) appear to be easier to value.

[Insert Figure 3 about here]

Graphs in Panel B of Figure 3 report the medians of valuation errors (a measure of bias) from machine learning models across firm-characteristic and industry portfolios. Machine learning models tend to overvalue small stocks but undervalue large stocks (Panel B.1 of Figure 3). Machine learning models also systematically overvalue stocks with low market-to-book ratios and undervalue those with high market-to-book ratios. By contrast, the machine learning model bias does not appear to vary substantially across ROE quintiles or industries.

4.3 Valuation error and future stock returns

Are machine valuations a good proxy for fundamental valuations? The challenge is that fundamental value is unobservable. High accuracy alone does not necessarily imply that a valuation is a good proxy for fundamental value, because market values might diverge from fundamental value in the short run. However, if a predicted valuation is a good proxy for fundamental value, then the market value of a stock should converge to the predicted value on average. Prices of expensive stocks (according to the model) should fall and prices of cheap stocks should rise. Valuation errors should predict future stock returns if machine valuations are good proxies for the fundamental values that market values converge to.

To test whether valuation errors predict future returns we turn to sorted portfolios. At the end of each month, we allocate stocks into quintile portfolios based on the ranks of valuation errors. Quintile one contains the most expensive stocks according to the model (e.g., stocks with the largest negative valuation errors) and quintile five contains the cheapest stocks. We implement a long-short strategy that buys the portfolio containing the cheapest stocks and shorts the portfolio consisting of the most expensive stocks for a month (quintile 5 minus

Electronic copy available at: https://ssrn.com/abstract=3740270

quintile 1). Valuation errors are estimated for all firms in each of the 480 months between January 1980 and December 2019. Valuation errors estimated in month t are used to form portfolios to be held in month $t+1$ (or, returns from February 1990 and December 2019).

Table 3 reports returns of value-weighted portfolios from trading strategies using valuation errors from all models, including the three machine learning models and all the traditional models. Table IA.7 in the Internet Appendix reports the returns of equal-weighted portfolios.[12]

[Insert Table 3 about here]

Trading strategies using machine valuation errors generate highly significant returns in both equal- and value-weighted portfolios. Across the board, equal-weighted strategies generate higher returns than value-weighted strategies. We focus on value-weighted strategies in our discussion because profits from equal-weighted strategies are difficult to realise due to their high turnover and transaction costs (Novy-Marx & Velikov, 2019). Monthly returns for value-weighted portfolios sorted on *ML* valuation errors range from 0.35% to 0.42% per month and are significant at a 1% level (shown as the estimates for *Const* in the panel tilted *Mean* in Table 3).

The trading profits from strategies utilising machine valuation errors are not simply compensation for bearing risk. The abnormal returns are slightly higher than the mean returns for value-weighted portfolios after we control for the Fama-French six-factor model (*FF6*) (Fama & French, 2018). The *FF6* alphas from value-weighted trading strategies using machine valuation errors range from 0.35% to 0.51% per month and are significant at a 1% level (shown as the estimates for *Const* in the panel titled *FF6* in Table 3).

To rule out the possibility that the discovered return predictability is a known anomaly, we use a 13-factor model, which is the *FF6* model augmented with seven "cluster portfolios" of anomaly-based trading strategies (shown as *FF6 + 7 clusters of anomalies* in Table 3 (value-weighted) and Table IA.7 in the Internet Appendix (equal-weighted)). Geertsema and Lu (2020) show that this augmented model renders insignificant 80 mean-significant anomalies previously identified in the literature. The return predictability from machine valuation errors is unlikely to be a prominent known anomaly, because the 13-factor model

[12]In order to mitigate the impact from microcap stocks, we follow Hou, Xue and Zhang (2018) in using NYSE breakpoints for both equal- and value-weighted portfolios and excluding firms below the 20th percentile of market capitalisation in each month when constructing equal-weighted portfolios.

Electronic copy available at: https://ssrn.com/abstract=3740270

(*FF6 + 7 clusters of anomalies*) only reduces the abnormal returns slightly and all abnormal returns remain highly significant.

Trading strategies using valuation errors from the *ML* models also generate higher Sharpe ratios than the trading strategies based on traditional models. Sharpe ratios of value-weighted *ML* strategies range between 0.44 and 0.54, while those from traditional strategies range between 0.10 and 0.40 (Table 3). The *ML* trading strategies also generate higher information ratios (0.55 to 0.80) than almost all target-matched traditional trading strategies.

4.4 Source of performance improvement

Does the superior performance of GBMs relative to traditional models come from using more information or from using a more powerful model? We find that using GBM models (instead of OLS or LASSO) and applying a log transformation to the target variable are the two most important sources of gains, while using a larger number of input variables is the least important source of gains.

Table 4 reconciles the performance of *BL* models (as a representative traditional model) to *ML* models. In Panel A of Table 4, we start with the *BL_m2b* model, where we fit an OLS model with the original 10 *BL* input variables to predict the market-to-book multiple (*m2b*). After we take a log transformation of the target variable (from *m2b* to *ln(m2b)*), the median absolute error (med($|\varepsilon|$)) drops by 0.12 from 0.52 to 0.40 and the dollar RMSE (rmse($\hat{M}$)) decreases by 3,813 from 12,437 to 8,624 million dollars while the out-of-sample R-squared ($R^2_{oos}(\hat{M})$) improves from 0.38 to 0.70. In the next step, we use a GBM model instead of an OLS regression. As a result, the median absolute error decreases by 0.07, the rmse($\hat{M}$) drops by 1,198 and the $R^2_{oos}(\hat{M})$ improves by 0.08. In these reconciliation steps above, all models rely on the variables used in the *BL model* (BL vars) and the same observations, so that any improvement in performance can be attributed to a change in methodology.

In the last step, we use all input variables available to the GBM model (ML vars). Using more variables does not materially improve the accuracy in predicting market-to-book multiples or value-to-asset multiples. By contrast, using more input variables substantially improves the accuracy in predicting value-to-sales multiples (from 0.39 to 0.32).

[Insert Table 4 about here]

Electronic copy available at: https://ssrn.com/abstract=3740270

Instead of GBM, one can also use LASSO to automatically select important input variables. Therefore we also perform a direct comparison between LASSO and GBM (see the last two rows in Panels A, B and C of Table 4). From LASSO to GBM, the reduction is 0.04 for median absolute error and 303 million dollars for the dollar RMSE when targeting $\ln(m2b)$. This result suggests that the outperformance of GBM is not only due to variable selection, since LASSO also performs variable selection. Instead, a large fraction of the power of GBMs appears to come from their ability to internally model non-linearities and interactions in the data, an ability which LASSO lacks.[13]

4.5 GBM peer weights

GBMs construct trees sequentially to improve prediction accuracy. Predicted valuation multiples are our GBM models' outputs. These outputs are analogous to the "warranted multiples" estimated from linear regressions as in Bhojraj and Lee (2002). In predicting valuation multiples, GBMs by-pass the step of selecting peer firms. To extract "peer firms" from GBM models, we develop a method to express GBM predictions as weighted averages of training data target multiples. In this setting, the peer universe comprises all firms in the training sample. If a training data firm is often allocated to the same leaf as the test firm in many trees, the training data firm will carry a higher peer weight. Conversely, if a training firm is rarely allocated to the same leaf as the test firm, the training firm will carry a lower peer weight. This GBM peer weight can be interpreted as the degree of comparability between a training data firm and the firm being valued. We detail the equations for calculating GBM peer weights in section A.3 in the Appendix.

As an illustration, Figure 4 depicts the distribution of peer weights of training firms in valuing Moderna in December 2019. The line in Figure 4 starts with an almost vertical section, suggesting that a small number of firms have high weights and hence are closely comparable to Moderna. By contrast, most firms have weights close to zero and a small number of firms have very negative weights, shown as a dip at the end of the line in Figure 4. That is, there are a small number of very dissimilar firms. These peer weights can be used to remove some of the subjectivity in the choice of peer firms.

[Insert Figure 4 about here]

[13] Gu, Kelly and Xiu (2020) also notes the importance of non-linearities and interactions in the context of risk premia prediction.

Electronic copy available at: https://ssrn.com/abstract=3740270

As an example, Table 5 reports the 10 most and 10 least comparable firms for Moderna, along with the logarithm of market-to-book multiples (*lnm2b*), peer weights, industries and the most important features in valuing Moderna. Eight out of the 10 most comparable companies belong to the Fama-French drug industry; in addition, they spend heavily on R&D, hold large cash balances, have low debt levels and generate very negative ROEs. Not all drug developers are deemed to be close peers by GBMs. For instance, two of the least comparable companies also belong to the drug industry. Unlike Moderna's close peers, Acorda Therapeutics has high levels of borrowing and do not keep as much cash or spend as heavily on R&D. Similarly, Catalyst Pharmaceutical spends much less on R&D compared to Moderna and has a positive ROE, unlike Moderna.

[Insert Table 5 about here]

5 Determinants of relative valuation

Does our machine learning approach construct models that are congruent with the underlying finance theory? This is an important question because GBMs take a data-driven approach. The important input variables chosen by GBMs broadly align with the theoretical determinants of fundamental value, suggesting that GBMs in a sense "learn" valuation theory from the data.

5.1 Theoretical determinants of valuation multiples

We start by constructing a simplified DCF model based on the Gordon constant growth model (Brennan, 1971; Gordon, 1962). For an equity valuation, we have

$$M_t = \frac{E_t[FCFE_{t+1}]}{r_e - g_e} = \frac{FCFE_t(1+g_e)}{r_e - g_e} \tag{1}$$

where $FCFE$ is the free cash flow to equity, r_e is firm's equity cost of capital and $g_e < r_e$ is the perpetual constant growth rate of the free cash flow to equity.

We express $FCFE$ as in Damodaran (2012) (p.352), where net capital expenditures and investment in net working capital are assumed to be financed with a fixed mix of debt and equity.

Electronic copy available at: https://ssrn.com/abstract=3740270

$$FCFE = NI - (CAPEX - DEPR)(1 - \delta) - \Delta NWC(1 - \delta) \tag{2}$$

NI is net income, $CAPEX$ is capital expenditures, $DEPR$ is depreciation, ΔNWC is investment in net working capital δ is the target debt ratio ($\frac{D}{D+E}$), or the proportion of net capital expenditures and investment in working capital that is financed by debt.

We define the equity market-to-book multiple as $m2b := \frac{M}{B}$ where M is the equity value of the firm and B is the book equity of the firm.

Replacing M in $m2b$ with equation (1) and then $FCFE$ with equation (2) yields

$$m2b = \left[\frac{NI}{B} - \frac{(CAPEX - DEPR) + \Delta NWC}{NI} \cdot \frac{NI}{B} \cdot (1 - \delta)\right] \cdot \frac{(1 + g_e)}{(r_e - g_e)} \tag{3}$$

$\frac{NI}{B}$ is the return on equity (ROE) while $\frac{(CAPEX-DEPR)+\Delta NWC}{NI}$ can be interpreted as the equity reinvestment rate (θ_e) (following Damodaran (2012), p. 353). We can now frame the market-to-book multiple in terms of the firm's return on equity (ROE), equity reinvestment rate (θ_e), debt ratio δ, growth rate g_e and cost of equity r_e.

$$m2b = ROE[1 - \theta_e(1 - \delta)]\frac{1 + g_e}{r_e - g_e} \tag{4}$$

Using a DuPont decomposition as in Soliman (2008), we can represent the same market-to-book multiple with a different set of accounting ratios. Incorporating sales (hereafter S) and assets (hereafter A) in equation (4), we construct $m2b$ with net profit margin (hereafter $PM1 := \frac{NI}{S}$), asset turnover ratio (hereafter $ATR := \frac{S}{A}$) and leverage (hereafter $LEV := \frac{A-B}{A}$) as below[14]

$$m2b = \frac{PM1 \cdot ATR}{1 - LEV}\left[1 - \theta_e(1 - \delta)\right]\frac{(1 + g_e)}{(r_e - g_e)} \tag{5}$$

Using a similar approach, the enterprise-value-to-assets multiple ($v2a := \frac{V}{A}$) can be constructed using return on assets (hereafter $ROA := (EBIT(1 - t))/A$), firm reinvestment

[14]Note $ROE = \frac{NI}{B} = (\frac{NI}{A}) \cdot (\frac{A}{B}) = (\frac{NI}{S} \cdot \frac{S}{A})/(\frac{1}{A} \cdot (A - (A - B))) = (\frac{NI}{S} \cdot \frac{S}{A})/(1 - (A - B)/A) = (PM1 \cdot ATR)/(1 - LEV)$.

Electronic copy available at: https://ssrn.com/abstract=3740270

rate (hereafter $\theta_v := \frac{(CAPEX-DEPR)+\Delta NWC}{EBIT(1-t)}$), growth in free cash flow to the firm (g) and weighted average cost of capital ($WACC$) as below

$$v2a = ROA(1-\theta_v) \cdot \frac{1+g}{WACC-g} \tag{6}$$

We can decompose return on assets (ROA) into an enterprise profit margin (hereafter $PM2 := \frac{EBIT(1-t)}{S}$) and asset turnover ($ATR$), or, $ROA = PM2 \cdot ATR$.

Then the enterprise-value-to-asset multiple ($v2a$) in equation (6) can be expressed with $PM2$ and ATR

$$v2a = PM2 \cdot ATR \cdot (1-\theta_v) \cdot \frac{1+g}{WACC-g} \tag{7}$$

Finally we consider the enterprise-value-to-sales multiple ($v2s$). By using $v2s = v2a \cdot \frac{A}{S}$ and equation (7) we can re-write $v2s$ in terms of enterprise profit margin ($PM2$)

$$v2s = v2a \cdot \frac{A}{S} = \frac{v2a}{ATR} = PM2 \cdot (1-\theta_v) \cdot \frac{1+g}{WACC-g} \tag{8}$$

Using operating leverage ($oplev := \frac{COGS+SGA}{A}$) and assuming $EBIT = S-(COGS+SGA)$, we can re-write enterprise profit margin ($PM2$) in terms of $oplev$ and ATR

$$PM2 = (1 - \frac{oplev}{ATR}) \cdot (1-t)$$

Substituting the above for $PM2$ in equation (8), we can define $v2s$ with asset turnover (ATR) and operating leverage ($oplev$)

$$v2s = (1 - \frac{oplev}{ATR}) \cdot (1-t) \cdot (1-\theta_v) \cdot \frac{1+g}{WACC-g} \tag{9}$$

Alternatively, we can frame $v2s$ in terms of return on assets (ROA) and asset turnover (ATR) by adapting equation (6).

$$v2s = v2a \cdot \frac{A}{S} = \frac{v2a}{ATR} = \frac{ROA}{ATR} \cdot (1-\theta_v) \cdot \frac{1+g}{WACC-g} \tag{10}$$

Electronic copy available at: https://ssrn.com/abstract=3740270

The relationship between the various valuation multiples and their theoretical determinants are summarised in Table 6.

[Insert Table 6 about here]

We infer expected marginal effects of an input variable on a valuation multiple from the theoretical relationships defined in equations (4) to (10).

Variables related to profitability (ROE, ROA and PM) and growth (g_e and g) are expected to have positive marginal effects on all valuation multiples because they directly drive future cash flows.

Some accounting ratios are more important for certain valuation multiples if they directly relate an accounting amount to profits. ROE and net profitability measures ($PM1$) should be more important for the market-to-book multiple (equations (4) and (5)), while ROA and EBIT margins ($PM2$) should be more important for enterprise-value multiples (equations (6) and (8)).

Efficiency ratios are expected to have a *positive* marginal effect for the enterprise-to-assets multiple (*v2a*) but *negative* for the enterprise-to-sales multiple (*v2s*). Equation (7) implies a positive marginal effect of asset turnover (ATR, the sales-to-assets ratio) on *v2a*. In contrast, ATR is negatively related to *v2s*, as is evident from equation (10). Another efficiency ratio, operating leverage (*oplev*), also has a theoretically negative marginal effect on *v2s* because enterprise profit margins and profits drop when operating leverage (*oplev*) increases, as per equation (9).

The impact of financial leverage (variables in the Financial soundness, Solvency and Liquidity categories) on all three valuation multiples can be positive or negative. For the market-to-book multiple ($m2b$), the positive effect works through the debt ratio (δ, see equations (4) and (5)). The negative effect operates through the discount rate channel (r_e, see equations (4) and (5)). For *v2a* and *v2s*, the marginal impact of financial leverage are also ambiguous because the direction of the impact on the weighted average cost of capital ($WACC$, see equations (7) to 10) depends on whether a firm's current leverage is above or below the optimal leverage (Modigliani & Miller, 1958).

Equations (4) to (10) cover all categories of input variables considered in this paper except the Size category and the Payout category. We expect that a firm's size has a negative marginal effect on valuation multiples, since small firms tend to have high expected growth

Electronic copy available at: https://ssrn.com/abstract=3740270

rates and hence high valuation multiples. If payout ratios are negatively related to firms' growth opportunities, the payout ratio should be negatively related to valuation multiples.

5.2 Determinants of valuation multiples identified by GBM

In OLS models the contribution of an independent variable is assessed by the magnitude and statistical significance of its coefficient. For GBM models we measure the contribution of a variable with SHAP values (SHapley Additive exPlanations).[15]

An input variable's SHAP value reflects the extent to which it moves the expected model output from the unconditional expectation (e.g. mean in the training data) towards the model prediction. SHAP values can be computed for each observation and are additive. For a given observation, the sum of SHAP values for all inputs added to the training sample mean produces the predicted target value.

[Insert Figure 5 about here]

Figure 5 illustrates the relationship between SHAP values and a prediction. The most important input in predicting the *lnm2b* multiple for Moderna in December 2019 is R&D (*rd_sale_wr*). It increases the predicted *lnm2b* from the unconditional training sample mean of 0.77 by 0.41 to 1.38. The Fama-French 49 industry classification is the second most important input, followed by the cash-to-assets ratio. The sum of SHAP values of all inputs is equal to the difference between the training sample mean and the predicted value of *lnm2b*. This additive feature also allows us to assess the average importance of a variable across an entire dataset or for a subset of data, as well as measuring the aggregate importance of groups of variables rather than individual variables.

[Insert Table 7 about here]

[15] SHAP values were introduced by Lundberg and Lee (2017), building on the Nobel Prize winning game theoretical approach of Shapley (1951) that measures the marginal contribution of each player in a cooperative game. The computational cost of calculating SHAP values quickly becomes infeasible for a large number of variables, since the number of calculations required grows exponentially in the number of variables evaluated. However, the recent TreeSHAP algorithm of Lundberg, Erion, Chen et al. (2020) enables the efficient calculation of exact SHAP values by exploiting regularities in tree-based models even for models with a large number of input variables. Interested readers can refer to Lundberg and Lee (2017) and Lundberg, Erion, Chen et al. (2020) for details regarding the computation of SHAP values.

Electronic copy available at: https://ssrn.com/abstract=3740270

Table 7 reports the 10 most important variables for each machine learning model based on the magnitude of their absolute percentage SHAP values (or, *SHAP values* or *variable importance* hereafter).[16] The SHAP values reported in Table 7 are averaged across all observations by variable for each model.

Most important variables in Table 7 are consistent with the theoretical drivers for fundamental valuation multiples. The most important accounting ratios in estimating market-to-book (*m2b*) multiples are ROE measures (*roe_wr*, or the ROE WRDS ratio and *profitability_an*, or profit before tax scaled by equity). In contrast, the two most important accounting return measures for enterprise-value-to-asset (*v2a*) multiples are return on invested capital (*roic_an*) and pre-tax return on total earning assets (*pretret_earnat_wr*), both of which are ROA-type measures. These findings conform to theoretical predictions because ROE measures directly relate book equity to profits (equation (4)) while ROA measures links assets to profits (equation (6)).

Operating leverage, asset turnover and before-interest margins connect sales to profits, and thus are important accounting ratios in determining the enterprise-value-to-sales multiple *v2s* (equations (9) and (10)). Based on SHAP values, the top three variables include two efficiency measures (*opleverage_an*, total cost scaled by total assets and *assetturnover_an*, asset turnover) and EBIT margin (*opmad_wr*).

Most correlations between the top 10 important variables and the target variable have the expected sign, as indicated by the actual rank correlation (*rho*) and the *Expected Sign* in Table 7.[17] Almost all profitability measures and all growth measures are positively correlated with target variables. Efficiency measures are positively correlated with the enterprise-value-to-assets multiple but negatively with the enterprise-value-to-sales multiple, consistent with theory. The only exception is *CAPMbeta*'s negative rank correlation with the target variable in the *ML_lnm2b* model, which contradicts the expected sign.[18]

To understand the importance of groups of accounting variables, we aggregate our variable importance results by variable *Category* in Table 8. The category importance largely conforms to the predicted relationships between accounting information and valuation formulas derived in section 5.1.

[16] We standardise SHAP values by scaling absolute SHAP values across the input variables for each observation so that their sum adds to 100%. All of SHAP results in this paper reports these scaled absolute SHAP values, unless otherwise indicated.

[17] We use rank correlation to be consistent with the tree-based method where samples are split based on a threshold value, which depends on the relative *rank* of the variable.

[18] Because the rank correlation (*rho*) is a univariate measure but theoretically expected signs are for the marginal effect (holding other variables constant), *rho* may not always have the theoretically expected sign.

Electronic copy available at: https://ssrn.com/abstract=3740270

[Insert Table 8 about here]

The Profitability category stands out because it contributes the most to the prediction of both the market-to-book (38.0%) and the value-to-assets (24.9%) valuation multiples and contributes the second most to value-to-sales valuation multiples (23.4%).

Growth measures make up the second most important category overall because it is ranked as the second most important category for both for the *ML_lnm2b* model (20.6%) and the *ML_lnv2a* model (17.4%) and the third most important category for the *ML_lnv2s* model (12.4%).

Efficiency ratios are the third most important category overall. This is the most important category in estimating value-to-sales multiples (36.5%), as well as making substantial contributions to predictions of market-to-book (6%) and value-to-assets (10.7%) multiples. This finding is notable because traditional relative valuation models do not explicitly include efficiency ratios. It mirrors the findings by Soliman (2008), which documents the value-relevance of asset turnover ratios.

Industry classification is not always the most important variable in predicting valuation multiples. As a single variable in Table 7, industry is the most important variable in market-to-book and value-to-asset valuations, but only the fourth most important variable in value-to-sales valuations. If we consider industry as a categorical variable that consists of 49 industry dummy variables, its importance is often ranked behind the Profitability, Growth, Efficiency and Financial Soundness categories in Table 8.

In short, the results in this subsection reinforce our emphasis on the importance of firm fundamentals.

5.3 Variable importance over time and across firms

The relationship between accounting information and firm value evolves over time. We observe three common themes. First, the Growth category in aggregate has become more important for valuation. Second, the Profitability category and the Efficiency category have always been important for valuation. Third, accounting ratios in the same category display diverging trends in importance. Ratios using book equity, after-tax profits, fixed-assets and cost of goods sold tend to decline in their importance for valuation. In contrast, ratios using R&D, pre-tax profits and current assets have become more important for valuation.

Electronic copy available at: https://ssrn.com/abstract=3740270

Overall, the variables related to Growth become more important over time for all three valuation multiples (Table 8). For example, SHAP values of the Growth category increase by 5.6 percentage points from 15.9% in the 1980s to 21.5% in the 2010s in predicting *lnm2b*. Although the aggregate importance of the Growth category has been rising, within the category the change in book equity (*chbe_an*) declined in importance in all three valuation models (Table 9). The driving force behind the rising importance of the Growth category is the R&D to Sales (*rd_sale_wr*) ratio, whose importance increase substantially in predicting the three target valuation multiples.

[insert Table 9 about here]

The importance of the Profitability category has remained largely unchanged for all three multiples over the last four decades. However, within this category after-tax profitability ratios (for example, Return on Equity or *roe_wr*) became less important while pre-tax and before-interest profitability measures (*profitability_an*, or profit before tax scaled by book equity) became more important in valuations (Panel *ML_lnm2b* of Table 9). This finding echoes Brown, Lo and Lys (1999), Collins, Maydew and Weiss (1997) and Francis and Schipper (1999) who show that the explanatory power of net income on stock price has declined over time.

For the enterprise-value-to-assets valuation multiple (*v2a*), Financial soundness ratios have become less relevant because their SHAP value decreased from 22.9% in the 1980s to 14.3% in the 2010s (Table 8). A closer examination (Table 9) reveals that the SHAP value of the Total Liability to PP&E ratio (*lt_ppent_wr*) decreases by 4.6 percentage points and that of the Accruals to Total Assets ratio (*Accrual_wr*) drops by 8.0 percentage points over the four decades.

While Financial soundness ratios have dropped in its importance for enterprise-value-to-assets (*v2a*) valuation, Efficiency ratios have become the second most important category for *v2a* in the 2010s. Its SHAP value has increased from 6.0% to 19.5% in the sample period (Table 8). This category has also remained as the most important category for predicting value-to-sales (*lnv2s*) multiples. Within this category, receivable turnover ratios (*sales2rec_an* and *rect_turn_wr*) and the asset turnover ratio (*at_turn_wr*) have become more important for valuation (Table 9). By contrast, the operating leverage ratio (*oplever-age_an*) declined substantially in importance; this result corroborates the finding in Barth, Li and McClure (2021) that cost of goods sold declined in value-relevance over time.

Electronic copy available at: https://ssrn.com/abstract=3740270

6 Conclusion

We use Gradient Boosting Machines (GBMs) to conduct relative valuation based on firm fundamentals. By using a machine learning approach, we provide new insights into relative valuation and the peer selection process. GBM models not only generate more accurate valuations than traditional models over time and across firms, their valuations also closely resemble fundamental value. Over-valued stocks (relative to the model) on average drop in price and under-valued stocks increase in price over the next month. We examine the relationship between accounting information and valuation multiples using SHAP values. Remarkably, the most influential variables in GBM predictions correspond closely to the variables derived from theoretical DCF models. An important benefit of GBM models is that they can identify complex relations between accounting information and firm value. Prior studies generally document a decrease in the importance of accounting information for equity valuation. Our results show that these declines relate mainly to the book value of equity, total costs, after-tax profits, PP&E and accruals. In a novel contribution to explainable AI we demonstrate that GBM predicted valuation multiples can be expressed as weighted averages of training sample valuation multiples. These GBM derived "peer weights" can be used for peer selection in valuation and research settings.

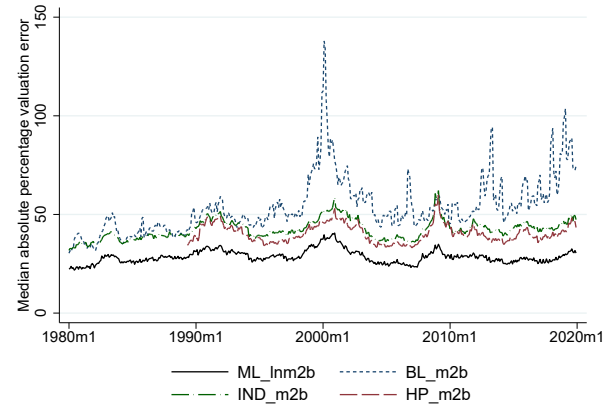
Our work demonstrates the potential of machine learning for conducting relative valuation and identifying peer groups. GBMs can assist researchers in making more objective choices of peers, as well as guiding researchers towards the most important firm fundamentals for the valuation task at hand. Our approach could be extended in several directions. A more sophisticated model with richer information (e.g. patent data, unstructured data from company filings or media announcements, etc.) might generate even more accurate predictions of valuation multiples and their associated peer weights. Even with advanced machine learning models, valuations still require substantial human judgement because many important drivers of valuations, such as network effects or a culture of innovation, are not readily observable in the historical data. This might explain why GBMs find small firms, low market-to-book firms and loss-making firms harder to value. However, as we show, machine learning excels at modelling complex patterns. In this way, machine learning can be used to complement human intuition and judgement.

Electronic copy available at: https://ssrn.com/abstract=3740270

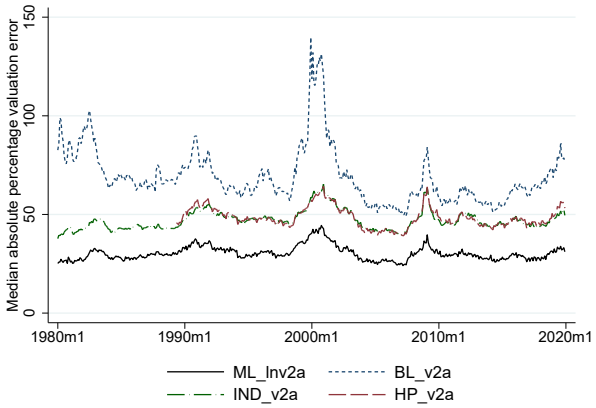
Figure 1: Comparison of model valuation errors over time

These figures depict medians of absolute valuation errors (med($|\varepsilon|$)) from the three base machine learning models and those from traditional models across time. Refer to Table 2 for model definitions. Panels A, B and C compare models targeting *m2b*, *v2a* and *v2s*, respectively. Panel D compares base machine learning models targeting *m2b*, *v2a* and *v2s* and those from *RRV* (targeting *lnm*) and *BG* (targeting *m*). Panel E compares the machine models. ε is the percentage valuation error: $\varepsilon := (\hat{M}/M) - 1$. $\hat{M}$ is the predicted equity value and M is the actual equity value. The sample period runs from January 1980 to December 2019.

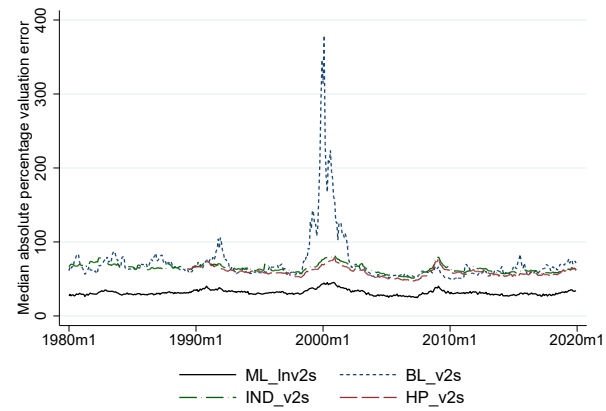
A: Models targeting *m2b*



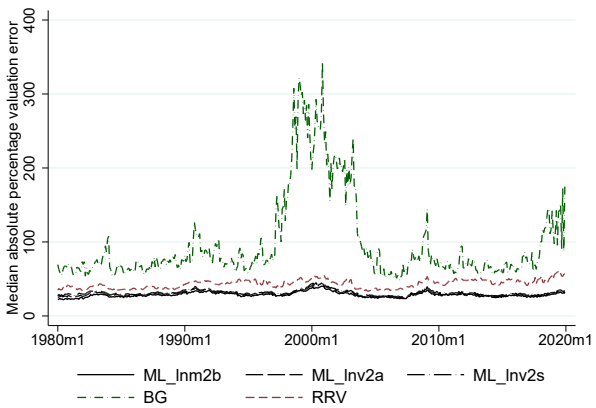
B: Models targeting *v2a*



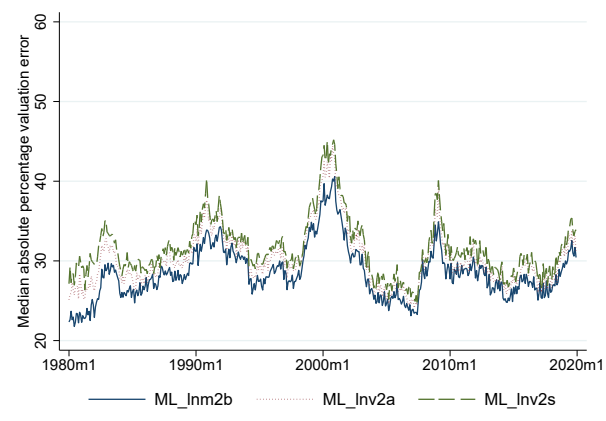
C: Models targeting *v2s*



D: *ML* models, *RRV* and *BG*



E: ML models only



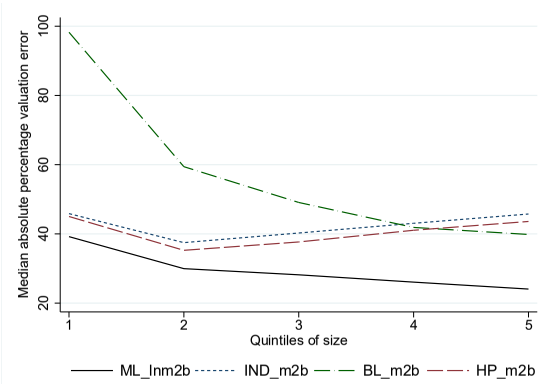
Electronic copy available at: https://ssrn.com/abstract=3740270

Figure 2: Valuation errors across firms

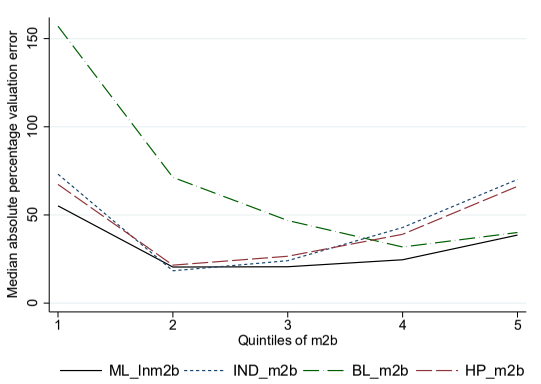
These figures compare medians of absolute valuation errors (med($|\varepsilon|$)) from the *ML_lnm2b* model to those from corresponding traditional models across firm characteristic quintiles. Firms are sorted into quintiles based on size, market-to-book ratio or ROE in each month. Graphs in Panel A compare models targeting *m2b*. Graphs in Panel B compare the *ML_lnm2b* model against those from *RRV* (targeting *lnm*) and *BG* (targeting *m*). Comparisons against the *ML_lnv2a* and *ML_lnv2s* models are in Figures IA.4 and IA.5 in the Internet Appendix. ε is the percentage valuation error: $\varepsilon := (\hat{M}/M) - 1$. $\hat{M}$ is the predicted equity value and M is the actual equity value. Table 2 provides model definitions. The sample period runs from January 1980 to December 2019.

A: Models targeting *m2b*

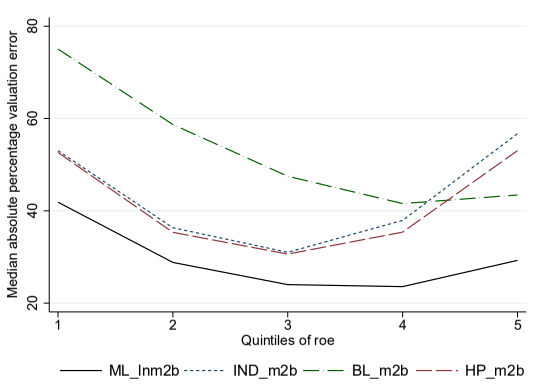
A.1: Size quintiles



A.2: Market-to-book quintiles

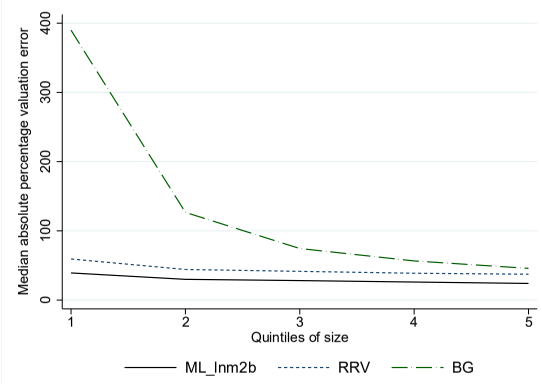


A.3: ROE quintiles

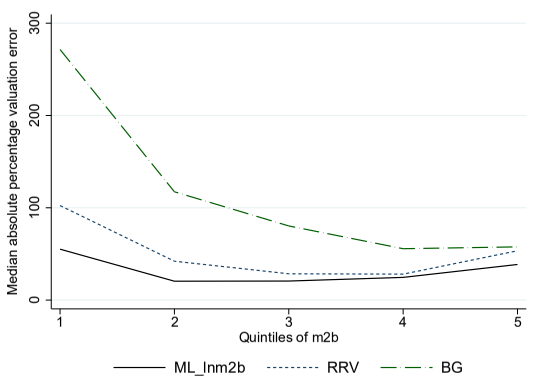


B: *ML_m2b*, *RRV* and *BG*

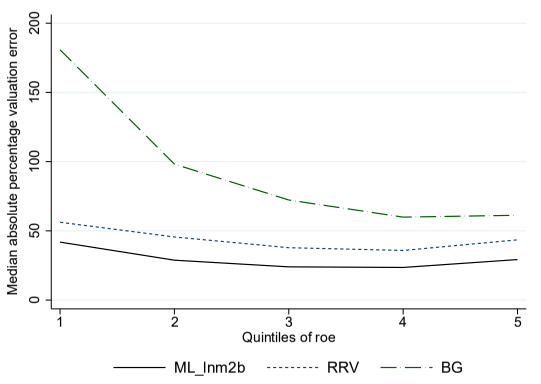
B.1: Size quintiles



B.2: Market-to-book quintiles



B.3: ROE quintiles



Electronic copy available at: https://ssrn.com/abstract=3740270

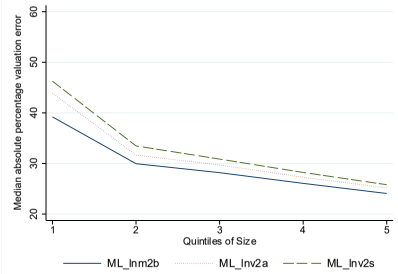
Electronic copy available at: https://ssrn.com/abstract=3740270

Figure 3: Valuation errors across firms for ML models

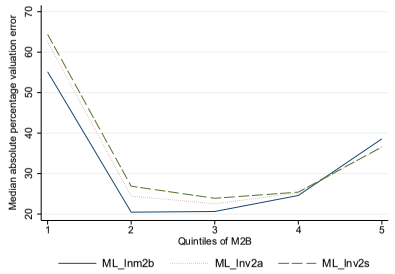
Figures in Panel A report medians of *absolute* valuation errors (med($|\varepsilon|$)) and figures in Panel B report medians of valuation errors (med(ε)). The comparison is across firm characteristic quintiles and Fama-French 12 industries for each of the three base machine learning models. Firms are sorted into quintiles based on size, market-to-book ratio or ROE in each month. Refer to Table 2 for model definitions. ε is the percentage valuation error (stated in percentage points): $\varepsilon := (\hat{M}/M) - 1$. $\hat{M}$ is the predicted equity value and M is the actual equity value. Table 2 provides model definitions. The sample period runs from January 1980 to December 2019.

A: Median of *absolute* percentage valuation errors (med($|\varepsilon|$))

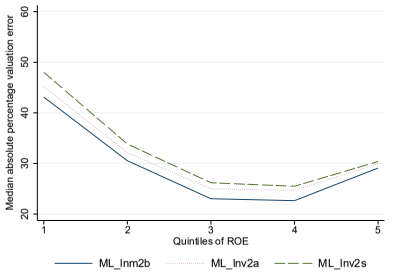
A.1: Size quintiles



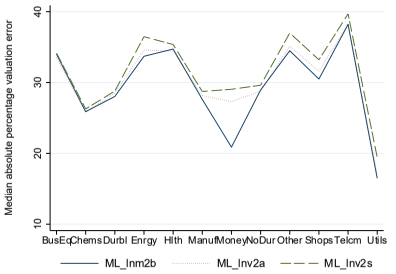
A.2: Market-to-book quintiles



A.3: ROE quintiles

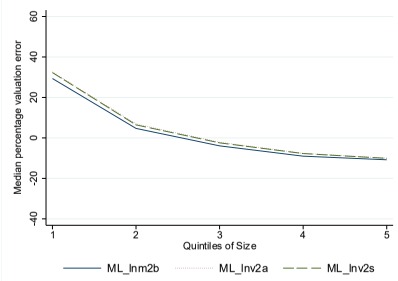


A.4: Across FF12 industries

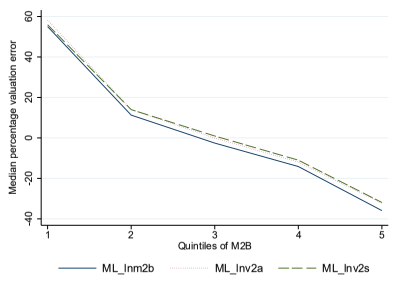

Median absolute percentage valuation error
BusEq Chems Durbl Enrgy Hlth Manuf Money NoDur Other Shops Telcm Utils
ML_lnm2b
ML_lnv2a
ML_lnv2s

B: Median of percentage valuation errors (med(ε))

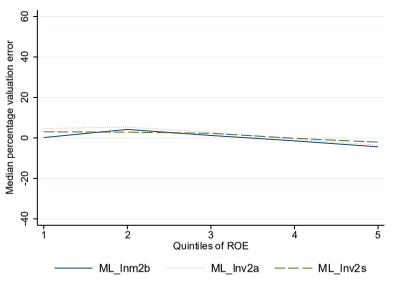
B.1: Size quintiles



B.2: Market-to-book quintiles



B.3: ROE quintiles



B.4: Across FF12 industries

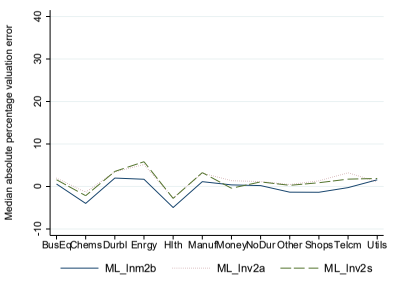
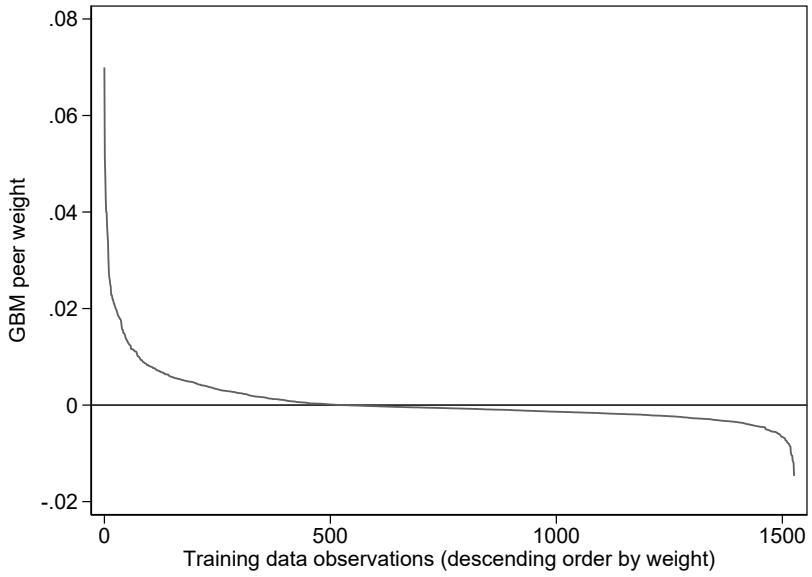

Median absolute percentage valuation error
BusEq Chems Durbl Enrgy Hlth Manuf Money NoDur Other Shops Telcm Utils
ML_lnm2b
ML_lnv2a
ML_lnv2s

Figure 4: Peer weights

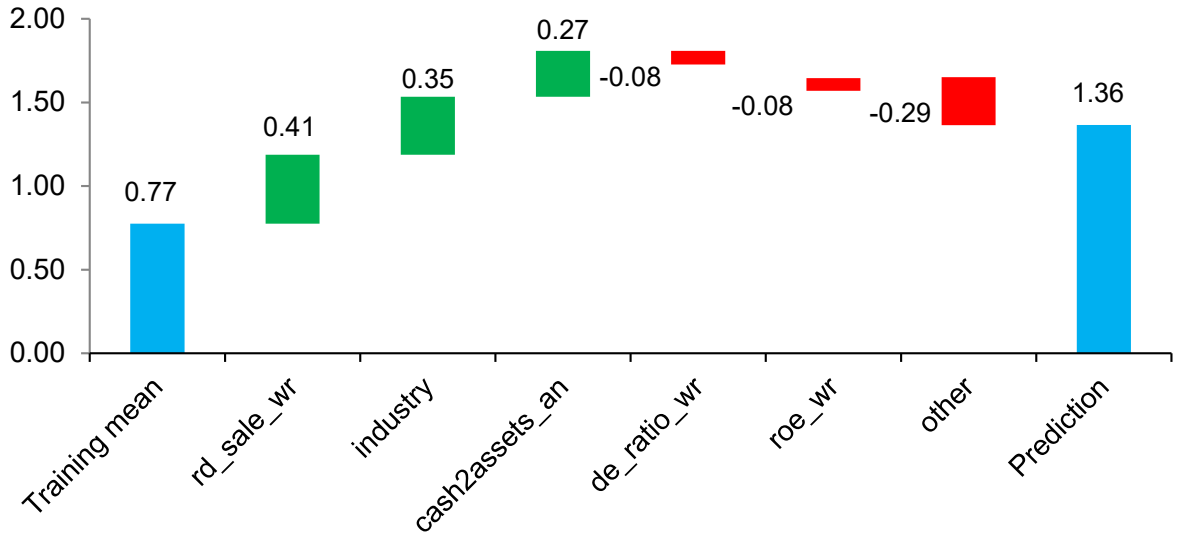
This figure plots the GBM peer weights of Moderna for December 2019. The GBM prediction of the log of market-to-book multiple (*lnm2b*) can be expressed as a weighted average of *lnm2b* target values of firms in the training data (refer to Section A.3 for details). Training data observations are sorted in descending order by GBM peer weight.



Electronic copy available at: https://ssrn.com/abstract=3740270

Figure 5: Individual SHAP values for Moderna (December 2019)

This figure shows the variables that contributed the most to predicting the log of the market-to-book multiple (*lnm2b*) of Moderna in December 2019. The unconditional prediction is the mean of *lnm2b* in the training sample ("Training mean"). "Prediction" is the out-of-sample prediction of *lnm2b* of Moderna in December 2019. We report the contribution from each of the five most important input variables. The contribution from other variables are aggregated into "other". The difference between the predicted *lnm2b* and the unconditional prediction is equal to the sum of the raw (not scaled absolute) SHAP values across all variables for a given observation.



Electronic copy available at: https://ssrn.com/abstract=3740270

Table 1: Machine learning data selection

This table reports the sample size after each step taken to obtain the final sample used by the machine learning models. We start with all US common equities (*shrcd* 10 or 11) with non-missing valuation multiples listed on NYSE, NASDAQ and AMEX (*exchcd* 1, 2 or 3) between January 1980 and December 2019 inclusive. We exclude firm-month observations with any of assets, book equity or sales below the 10th percentile in each month in order to mitigate the impact from tiny firms.

Steps	Firm-Months	Firms
Common domestic stocks listed on NYSE, NASDAQ or AMEX	2,361,860	19,697
Positive and non-missing targets (m2b, v2a, v2s)	2,162,501	19,041
All scale variables above 10th percentile in each month	1,811,786	16,201

Electronic copy available at: https://ssrn.com/abstract=3740270

Electronic copy available at: https://ssrn.com/abstract=3740270

Table 2: Model performance comparison

This table reports summary statistics of valuation errors (ε), root mean squared errors ($rmse$), out-of-sample R-squared's (R^2_{oos}) and Pearson correlations ($\rho_{M,\hat{M}}$) from different models. Row headings are valuation models. *ML_lnm2b*, *ML_lnv2a* and *ML_lnv2s* refer to three *base machine learning models* using *lnm2b* (the log of market-to-book multiple), *lnv2a* (the log of enterprise-value-to-assets multiple) and *lnv2s* (the log of enterprise-value-to-sales) as target variables, respectively. *BL_m2b*, *BL_v2a* and *BL_v2s* refer to monthly cross-sectional regression models based on Bhojraj and Lee (2002) using valuation multiples as target variables. *IND_m2b*, *IND_v2a* and *IND_v2s* refer to the industry harmonic mean approach in Liu, Nissim and Thomas (2002). *HP_m2b*, *HP_v2a* and *HP_v2s* refer to the peer group mean approach based on Hoberg and Phillips (2010, 2016) where a firm's peer group consists of 60 firms (minimum 5) with the most similar business descriptions in last year's 10-k filings (https://hobergphillips.tuck.dartmouth.edu/industryclass.htm). *RRV* refers to an industry-month cross-sectional model used to estimate equity value based on Rhodes-Kropf, Robinson and Viswanathan (2005). *BG* refers to a monthly cross-sectional regression approach to value equity based on Bartram and Grinblatt (2018). Section 2 contains additional valuation model details. N is the number of observations. $\hat{M}$ is the predicted equity value and M is the actual equity value, both in millions of dollars. ε is the percentage valuation error: $\varepsilon := (\hat{M}/M) - 1$. med$(\cdot)$, avg$(\cdot)$ and sd$(\cdot)$ denote the median, mean and standard deviation, respectively. $rmse(\hat{x}) := \sqrt{\frac{1}{N}\sum(x-\hat{x})^2}$. $rmse(\hat{y})$ is the *model* RMSE for target variable y and $rmse(\hat{M})$ is the *dollar* RMSE for equity value M. $R^2_{oos}(\hat{p}) := 1 - \Sigma(p-\hat{p})^2/\Sigma(p-\bar{p})^2$. $R^2_{oos}(\hat{y})$ and $R^2_{oos}(\hat{M})$ are the model and dollar out-of-sample R-squared measures, each based on the target variable and the dollar equity value respectively.

	N	med($\lvert\varepsilon\rvert$)	avg($\lvert\varepsilon\rvert$)	med(ε)	avg(ε)	sd(ε)	rmse($\hat{y}$)	rmse($\hat{M}$)	$r^2_{oos}(\hat{y})$	$r^2_{oos}(\hat{M})$	$\rho_{M,\hat{M}}$
ML_lnm2b	1,811,785	0.29	0.48	-0.00	0.19	1.14	0.56	7,289	0.54	0.78	0.90
ML_lnv2a	1,811,785	0.30	0.54	0.01	0.25	1.53	0.42	7,395	0.75	0.77	0.89
ML_lnv2s	1,811,785	0.32	0.59	0.01	0.28	4.66	0.44	7,537	0.80	0.76	0.89
BL_m2b	1,335,453	0.52	1.16	0.21	0.73	3.91	6.47	12,437	0.12	0.38	0.79
BL_v2a	1,335,453	0.67	1.52	0.60	1.30	4.41	2.00	11,414	0.20	0.48	0.86
BL_v2s	1,335,453	0.68	2.25	0.20	0.85	14.13	7.29	12,245	0.21	0.40	0.83
IND_m2b	1,738,553	0.42	0.58	-0.20	0.02	1.32	6.58	12,200	-0.03	0.38	0.64
IND_v2a	1,738,553	0.47	0.86	-0.15	0.12	6.06	2.13	13,029	0.02	0.29	0.61
IND_v2s	1,738,553	0.64	1.29	-0.35	0.00	12.89	8.10	17,078	-0.01	-0.22	0.38
HP_m2b	905,568	0.40	0.61	-0.16	0.10	1.80	6.21	13,079	-0.03	0.51	0.73
HP_v2a	905,566	0.48	0.93	-0.17	0.04	7.74	1.84	16,681	0.09	0.20	0.60
HP_v2s	905,573	0.59	1.28	-0.22	0.17	9.55	9.64	30,563	-1.57	-1.67	0.28
RRV	1,576,212	0.43	0.73	0.02	0.35	2.07	0.76	12,050	0.83	0.41	0.65
BG	1,466,704	0.86	3.84	-0.17	-1.58	32.05	24186.12	24,186	-1.35	-1.35	0.42

Electronic copy available at: https://ssrn.com/abstract=3740270

Table 3: Value-weighted portfolio returns

This table reports monthly means and abnormal returns of value-weighted hedge portfolio returns from trading strategies utilising valuation errors. At the end of each month, stocks are sorted into quintiles by valuation error based on NYSE break points. One-month-ahead returns in each portfolio are weighted by the current month equity value. The strategy takes short positions in the quintile 1 (overvalued stocks) and long positions in quintile 5 (undervalued stocks) at the beginning of the next month, which is then held for one month. We use two asset-pricing models to estimate abnormal returns. One is the Fama-French six-factor (FF6) model. The other is the FF6 model augmented by seven additional factors which are shown to subsume all mean-significant anomalies accumulated to date by Geertsema and Lu (2020). Refer to Table 2 for a summary of the valuation models. Means and abnormal returns are in percentage points. Sharpe ratios and information ratios are annualised. The sample runs from January 1980 to December 2019 with 479 monthly returns in trading strategies. Coefficients significant at 10%, 5% and 1% from two-tailed t-tests are indicated with *, ** and ***. t-Statistics are based on Newey-West standard errors with a lag of three months.

	ML_lnm2b	ML_lnv2a	ML_lnv2s	BL_m2b	BL_v2a	BL_v2s	IND_m2b	IND_v2a	IND_v2s	HP_m2b	HP_v2a	HP_v2s	RRV	BG
Mean														
Const	0.37***	0.42***	0.35***	0.29*	0.32*	0.30**	0.18	0.16	0.31**	0.10	0.09	0.38**	0.33**	0.28**
Sharpe Ratio	0.44	0.54	0.44	0.31	0.28	0.36	0.18	0.18	0.33	0.10	0.10	0.40	0.37	0.35
FF6														
MKTRF	-0.03	-0.01	0.05**	-0.05	-0.00	0.13***	0.01	0.07***	-0.02	-0.00	-0.01	-0.11***	0.10***	0.04
SMB	0.09**	0.04	0.05	0.15***	0.15***	0.06	0.13***	0.10***	0.14***	0.14***	0.16***	0.22***	0.47***	-0.20***
HML	0.32***	0.21***	0.29***	0.24***	0.58***	0.01	0.59***	0.48***	-0.14**	0.41***	0.19***	-0.02	0.40***	0.27***
RMW	0.07	0.01	0.07	-0.17***	-0.05	0.06	-0.23***	-0.01	0.58***	-0.18***	0.06	0.25***	-0.00	0.28***
CMA	-0.08	0.04	0.08	-0.08	-0.02	0.09	-0.07	0.13*	0.57***	0.18**	0.41***	0.60***	0.05	0.05
UMD	-0.30***	-0.27***	-0.27***	-0.30***	-0.42***	-0.12***	-0.30***	-0.26***	-0.11***	-0.27***	-0.24***	-0.16***	-0.21***	-0.16***
Const	0.46***	0.51***	0.35***	0.51***	0.43***	0.23*	0.30***	0.11	0.05	0.21	0.08	0.31**	0.23**	0.18
Adj. R^2	0.36	0.30	0.36	0.28	0.52	0.07	0.55	0.50	0.25	0.44	0.35	0.31	0.52	0.35
Information Ratio	0.71	0.80	0.55	0.66	0.58	0.31	0.48	0.17	0.07	0.31	0.12	0.43	0.39	0.29
FF6 + 7 clusters of anomalies														
MKTRF	-0.05*	-0.04	0.02	-0.06*	-0.01	0.12***	-0.01	0.06**	-0.05	-0.02	-0.01	-0.09**	0.10***	0.04
SMB	0.07*	0.03	0.03	0.14***	0.12**	0.06	0.11***	0.08**	0.12**	0.07	0.09*	0.18***	0.42***	-0.22***
HML	0.34***	0.23***	0.31***	0.27***	0.59***	0.02	0.60***	0.48***	-0.13**	0.42***	0.18***	-0.02	0.41***	0.28***
RMW	-0.04	-0.09	-0.02	-0.29***	-0.16**	0.01	-0.27***	-0.03	0.53***	-0.28***	-0.02	0.13	-0.12**	0.24***
CMA	-0.07	0.06	0.08	-0.11	-0.01	0.09	-0.08	0.12	0.58***	0.14	0.34***	0.56***	0.03	0.07
UMD	-0.28***	-0.26***	-0.23***	-0.29***	-0.41***	-0.13***	-0.28***	-0.24***	-0.09***	-0.25***	-0.19***	-0.15***	-0.19***	-0.17***
EG	0.13*	0.11	0.11	0.23***	0.06	0.04	0.04	-0.02	-0.01	-0.01	-0.06	0.03	0.05	-0.05
Accruals	-0.17***	-0.17***	-0.17***	-0.16***	-0.15***	-0.11**	-0.09**	-0.10**	-0.20***	-0.12**	-0.23***	-0.26***	-0.11***	-0.07
STReversal	0.15***	0.12***	0.16***	0.10***	0.08**	0.01	0.09***	0.04*	0.04	0.10***	0.04	-0.01	0.08***	-0.00
MarginGrowth	0.10**	0.08	0.04	0.11**	0.03	0.05	0.04	0.00	0.04	0.08	-0.10*	-0.01	-0.02	0.06
epsconsistency	-0.04	-0.06**	-0.07**	-0.08**	0.04	-0.04	-0.05	-0.06*	-0.12***	0.06	0.01	0.01	0.10***	0.03
ExternalFinance	0.03	-0.02	0.03	-0.03	0.01	0.08	0.04	0.03	0.03	0.08	0.10*	0.06	0.01	-0.08*
seasonality	0.05*	0.08***	0.07***	0.02	0.02	0.08**	0.02	0.03	0.11***	-0.06*	0.01	0.03	-0.05*	-0.02
Const	0.38***	0.46***	0.29**	0.42***	0.45***	0.21	0.28**	0.16	0.13	0.28**	0.25*	0.43***	0.26**	0.29**
Adj. R^2	0.42	0.36	0.43	0.31	0.53	0.08	0.57	0.51	0.28	0.47	0.39	0.34	0.55	0.36
Information Ratio	0.62	0.76	0.49	0.56	0.61	0.28	0.46	0.27	0.17	0.44	0.39	0.61	0.46	0.47

Table 4: Source of performance gains

This table reports summary statistics of valuation errors (ε), root mean squared errors ($rmse$), out-of-sample R-squared's (R^2_{oos}) and Pearson correlations of actual with predicted equity values ($\rho_{M,\hat{M}}$) of OLS models, LASSO models and GBMs. The row headings indicate the methodology, the input variables and the target variable. For example, "OLS : BL vars → m2b" indicates that the results are from an OLS model using the variables in the *BL* model as inputs to predict market-to-book multiples. "GBM : ML vars (nm) → ln(m2b)" indicates that the results are from a GBM model using all observations with non-missing values for machine learning variables to predict the logs of market-to-book multiples. Refer to Section 2 for valuation model details. N is the number of observations.

A: Performance gains with m2b multiples

	N	med($\lvert\varepsilon\rvert$)	avg($\lvert\varepsilon\rvert$)	rmse($\hat{y}$)	rmse($\hat{M}$)	$r^2_{oos}(\hat{y})$	$r^2_{oos}(\hat{M})$	$\rho_{M,\hat{M}}$
OLS : BL vars → m2b	1,335,453	0.52	1.16	6.47	12,437	0.12	0.38	0.79
OLS : BL vars → ln(m2b)	1,335,453	0.40	0.67	0.71	8,624	0.30	0.70	0.85
GBM : BL vars → ln(m2b)	1,335,453	0.33	0.56	0.63	7,426	0.46	0.78	0.89
GBM : ML vars → ln(m2b)	1,335,453	0.31	0.50	0.58	7,632	0.54	0.77	0.89
LASSO : ML vars (nm) → ln(m2b)	273,225	0.30	0.45	0.51	14,200	0.51	0.76	0.89
GBM : ML vars (nm) → ln(m2b)	273,225	0.26	0.39	0.46	13,897	0.63	0.78	0.90

B: Performance gains with v2a multiples

	N	med($\lvert\varepsilon\rvert$)	avg($\lvert\varepsilon\rvert$)	rmse($\hat{y}$)	rmse($\hat{M}$)	$r^2_{oos}(\hat{y})$	$r^2_{oos}(\hat{M})$	$\rho_{M,\hat{M}}$
OLS : BL vars → v2a	1,335,453	0.67	1.52	2.00	11,414	0.20	0.48	0.86
OLS : BL vars → ln(v2a)	1,335,453	0.56	1.29	0.54	9,585	0.33	0.63	0.86
GBM : BL vars → ln(v2a)	1,335,453	0.34	0.61	0.48	6,721	0.48	0.82	0.91
GBM : ML vars → ln(v2a)	1,335,453	0.31	0.54	0.44	7,296	0.61	0.79	0.90
LASSO : ML vars (nm) → ln(v2a)	273,225	0.40	0.81	0.39	13,375	0.47	0.79	0.89
GBM : ML vars (nm) → ln(v2a)	273,225	0.26	0.40	0.34	13,179	0.67	0.80	0.91

C: Performance gains with v2s multiples

	N	med($\lvert\varepsilon\rvert$)	avg($\lvert\varepsilon\rvert$)	rmse($\hat{y}$)	rmse($\hat{M}$)	$r^2_{oos}(\hat{y})$	$r^2_{oos}(\hat{M})$	$\rho_{M,\hat{M}}$
OLS : BL vars → v2s	1,335,453	0.68	2.25	7.29	12,245	0.21	0.40	0.83
OLS : BL vars → ln(v2s)	1,335,453	0.54	1.22	0.73	12,756	0.46	0.35	0.78
GBM : BL vars → ln(v2s)	1,335,453	0.39	0.72	0.54	7,713	0.70	0.76	0.89
GBM : ML vars → ln(v2s)	1,335,453	0.32	0.56	0.45	7,165	0.79	0.79	0.91
LASSO : ML vars (nm) → ln(v2s)	273,225	0.32	0.51	0.41	15,511	0.73	0.72	0.85
GBM : ML vars (nm) → ln(v2s)	273,225	0.26	0.41	0.35	12,934	0.84	0.81	0.92

Electronic copy available at: https://ssrn.com/abstract=3740270

Table 5: Most and least comparable peers of Moderna

This table presents the 10 most and the 10 least comparable peers of Moderna for December 2019. The most (least) comparable peers are firms with the highest (lowest) weights, thus can be viewed as the most (least) comparable peers. Moderna in December 2019 had an actual *lnm2b* of 1.66 and a predicted *lnm2b* of 1.36.

company	lnm2b	weight	cumweight	ff12	ff49	rd_sale_wr	cash2assets_an	de_ratio_wr	roe_wr
Test firm									
Moderna Inc	1.66	n/a	n/a	Hlth	DRUGS	6.47	0.67	0.27	-0.42
Most comparable peers									
bluebird bio Inc	1.19	0.07	0.07	Hlth	DRUGS	10.02	0.62	0.29	-0.49
Guardant Health Inc	2.20	0.05	0.12	Hlth	DRUGS	0.42	0.55	0.11	-0.09
Amicus Therapeutics Inc	1.50	0.05	0.17	Hlth	DRUGS	0.77	0.58	0.61	-0.62
Liveramp Holdings Inc	1.01	0.04	0.21	BusEq	SOFTW	0.31	0.58	0.15	-0.12
Sangamo Therapeutics Inc	0.82	0.04	0.25	Hlth	DRUGS	1.89	0.56	0.53	-0.28
MacroGenics Inc	0.74	0.04	0.29	Hlth	DRUGS	3.47	0.72	0.38	-0.65
Agios Pharmaceuticals Inc	1.99	0.04	0.33	Hlth	DRUGS	3.54	0.65	0.56	-0.89
BridgeBio Pharma Inc	2.24	0.04	0.36	Hlth	DRUGS	7.62	0.84	0.25	-0.49
Insmed Inc	1.95	0.03	0.40	Hlth	DRUGS	1.38	0.70	1.51	-0.96
TiVo Corp	-0.21	0.03	0.43	BusEq	SOFTW	0.24	0.11	0.96	-0.36
Least comparable peers									
Cardlytics Inc	2.57	-0.01	0.98	BusEq	SOFTW	0.07	0.49	0.52	-0.25
Phreesia Inc	2.23	-0.01	1.00	BusEq	SOFTW	0.15	0.61	0.58	-0.23
Inspire Medical Systems Inc	2.50	-0.01	1.01	Hlth	MEDEQ	0.16	0.83	0.25	-0.20
Cryoport Inc	1.72	-0.01	1.02	Hlth	MEDEQ	0.11	0.69	0.22	-0.18
Roku Inc	3.30	-0.01	1.03	Durbl	TOYS	0.24	0.43	0.93	-0.08
Tandem Diabetes Care Inc	3.05	-0.01	1.04	Hlth	MEDEQ	0.13	0.54	0.75	-0.14
Q2 Holdings Inc	2.32	-0.01	1.05	BusEq	SOFTW	0.24	0.65	1.56	-0.18
Acorda Therapeutics Inc	-1.07	-0.01	1.06	Hlth	DRUGS	0.37	0.29	2.13	-1.16
Catalyst Pharmaceutical Inc	1.60	-0.01	1.07	Hlth	DRUGS	0.29	0.79	0.24	0.12
Bandwidth Inc	1.48	-0.01	1.08	BusEq	SOFTW	0.14	0.60	0.19	0.01

Electronic copy available at: https://ssrn.com/abstract=3740270

Table 6: Theoretical determinants of valuation multiples

This table summarises the theoretical determinants of valuation multiples. All equations are derived from simplified discounted cash flow valuation models as shown in Section 5.1, which decomposes valuation multiples into accounting ratios (see heading *Equation* in the table). *Category* refers to the category a variable belongs to in the machine learning input variable set. There are 11 categories in total, primarily based on the WRDS ratio suite classification. Variable definitions can be found in section A.1 of the Appendix. Section IA.2 of the Internet Appendix describes variable categories and reports summary statistics.

Multiple	Equation	Eq	Variable	Category	Expected Sign
$m2b$	$m2b = \frac{M}{B} = ROE[1-\theta_e(1-\delta)] \cdot \frac{1+g_e}{r_e-g_e}$	(4)	ROE	Profitability	+
			g_e	Growth	+
	$m2b = \frac{PM1 \cdot ATR}{1-LEV}\,[1-\theta_e \cdot (1-\delta)] \cdot \frac{(1+g_e)}{(r_e-g_e)}$	(5)	$PM1$	Profitability	+
			ATR	Efficiency	+
			LEV	Financial soundness/Solvency	−
	$\delta = \frac{D}{D+E}$		$\frac{D}{D+E}$	Financial soundness/Solvency	+
	$r_e = r^f + \beta_e Mktrf$	Capital-asset pricing model (CAPM)	β_e	CAPMbeta	−
	$\beta_e = \beta_u(1+\frac{D}{E})$;	Modiligani and Miller Theory	$\frac{D}{E}$	Financial soundness/Solvency	−
$v2a$	$v2a = \frac{V}{A} = ROA(1-\theta_v)\frac{1+g}{WACC-g}$	(6)	ROA	Profitability	+
	$v2a = \frac{V}{A} = PM2 \cdot ATR(1-\theta_v)\frac{1+g}{WACC-g}$	(7)	$PM2$	Profitability	+
			ATR	Efficiency	+
			g	Growth	+
	$WACC = \frac{E}{D+E}r_e + \frac{D}{D+E}r_d(1-t)$	Weighted average cost of capital	$\frac{D}{D+E}; r_d$	Financial soundness/Solvency	+/−
$v2s$	$v2s = \frac{V}{S} = PM2(1-\theta_v) \cdot \frac{1+g}{WACC-g}$	(8)	$PM2$	Profitability	+
	$v2s = (1-\frac{oplev}{ATR}) \cdot (1-t) \cdot (1-\theta_v) \cdot \frac{1+g}{WACC-g}$	(9)	$oplev$	Efficiency	−
	$v2s = \frac{V}{S} = \frac{ROA}{ATR}(1-\theta_v) \cdot \frac{1+g}{WACC-g}$	(10)	ROA	Profitability	+
			ATR	Efficiency	−
			g	Growth	+
	$WACC = \frac{E}{D+E}r_e + \frac{D}{D+E}r_d(1-t)$	Weighted average cost of capital	$\frac{D}{D+E}; r_d$	Financial soundness/Solvency	+/−

Electronic copy available at: https://ssrn.com/abstract=3740270

Table 7: Variable importance

This table presents the 10 most important variables as measured by the *SHAP* values for each machine learning model. *SHAP* is the absolute percentage SHAP values in percentage points. The *SHAP* measure for an input variable is the *SHAP* value averaged across all observations. We also report variable category (*Category*), cumulative SHAP (*CSHAP*) and the correlation of rank percentiles between each variable with the predicted target variable (*rho*). There are 11 categories in total, primarily based on the WRDS ratio suite classification. Variable definitions can be found in section A.1 of the Appendix. Section IA.2 of the Internet Appendix describes variable categories and reports summary statistics.

Variable	Description	Category	SHAP	CSHAP	rho	Expected sign
ML_lnm2b						
industry	Fama French 49 industries	Industry	8.9	8.9	-0.139	N.A.
roe_wr	Return on Equity	Profitability	7.7	16.6	0.536	+
profitability_an	Profitability (pbt/equity)	Profitability	6.0	22.6	0.490	+
rd_sale_wr	R&D/Sales	Growth	4.9	27.5	0.269	+
salesgrowth_an	Sales Growth	Growth	3.8	31.3	0.368	+
aftret_invcapx_wr	After-tax Return on Invested Capital	Profitability	3.7	34.9	0.554	+
pretret_noa_wr	Pre-tax Return on Net Operating Assets	Profitability	3.0	38.0	0.493	+
roic_an	Return on Invested Capital	Profitability	2.9	40.8	0.518	+
roce_wr	Return on Capital Employed	Profitability	2.4	43.3	0.517	+
CAPMbeta	CAPM beta	CAPMbeta	2.1	45.3	0.170	-
ML_lnv2a						
industry	Fama French 49 industries	Industry	7.7	7.7	-0.335	N.A.
roic_an	Return on Invested Capital	Profitability	7.0	14.7	0.535	+
chbe_an	Δ Book Equity	Growth	4.3	19.0	0.491	+
lt_ppent_wr	Total Liabilities/PP&E	Financial soundness	4.1	23.1	-0.353	+/-
debt_at_wr	Long-term Debt/Total Assets	Solvency	4.1	27.2	-0.121	+/-
sales2rec_an	Sales/Receivables	Efficiency	3.9	31.1	0.295	+
rd_sale_wr	R&D/Sales	Growth	3.8	34.9	0.352	+
pretret_earnat_wr	Pre-tax Return on Total Earning Assets	Profitability	3.3	38.2	0.494	+
salesgrowth_an	Sales Growth	Growth	3.2	41.4	0.319	+
lt_debt_wr	Long-term Debt/Total Liabilities	Financial soundness	2.9	44.3	0.016	+/-
ML_lnv2s						
opleverage_an	Operating Leverage	Efficiency	21.2	21.2	-0.819	-
at_turn_wr	Asset Turnover	Efficiency	9.0	30.2	-0.800	-
opmad_wr	EBIT Margin	Profitability	6.4	36.7	0.533	+
industry	Fama French 49 industries	Industry	5.9	42.6	0.112	N.A.
npm_wr	Net Profit Margin	Profitability	3.4	46.0	0.455	+
rd_sale_wr	R&D/Sales	Growth	2.5	48.5	0.072	+
salesgrowth_an	Sales Growth	Growth	2.4	50.8	0.191	+
debt_at_wr	Long-term Debt/Total Assets	Solvency	2.1	53.0	-0.045	+/-
gpm_wr	Gross Profit Margin	Profitability	2.1	55.1	0.539	+
chbe_an	Δ Book Equity	Growth	2.1	57.2	0.166	+

Electronic copy available at: https://ssrn.com/abstract=3740270

Table 8: Category importance

This table presents variable category importance as measured by the magnitude of *SHAP* values over the full sample and in each decade starting 1980, 1990, 2000 and 2010. Category *SHAP* is the sum of the *SHAP* values of all input variables included in a category. Each row in a table represents a variable category. There are 11 categories in total, primarily based on the WRDS ratio suite classification. Variable definitions can be found in Appendix A.1. Internet Appendix IA.2 describes variable categories and reports summary statistics. *rho* is the correlation between the average of rank percentiles of variables in the category and the rank percentiles of the predicted target variable. We re-orient the rank of a variable as needed so that it conforms to the semantics of the category label. For example, in calculating the average rank of inputs for the financial soundness category, we re-orient the rank for the debt-to-equity ratio (*dltt_be_wr*) so that a firm with *lower* debt-to-equity ratio is ranked *higher*, because it is more financially sound than firms with higher debt-to-equity ratios. *Expected sign* is the sign of the theoretical marginal impact from a category of variables on the target variable derived from the equations in Table 6. *Chg* is the change in SHAP value from the 1980 decade to the 2010 decade.

Category	SHAP	rho	Expected sign	1980	1990	2000	2010	Chg
MLlnm2b								
Profitability	38.0	0.648	+	38.6	39.1	36.6	37.6	-1.1
Growth	20.6	0.432	+	15.9	22.9	21.3	21.5	5.6
Industry	8.9	-0.123	N.A.	11.1	8.7	8.3	7.4	-3.7
Financial soundness	8.7	0.184	+/-	9.1	7.8	8.6	9.9	0.8
Efficiency	6.0	0.228	+	6.4	5.6	6.1	6.0	-0.4
Liquidity	5.7	0.198	+/-	5.2	4.4	7.0	6.4	1.2
Solvency	4.6	0.334	+/-	4.1	3.7	5.4	5.7	1.6
Size	3.7	0.024	-	4.4	4.2	3.3	2.3	-2.1
CAPMbeta	2.1	0.142	-	3.4	1.8	1.8	1.3	-2.2
Capitalisation	1.1	0.119	+/-	1.2	1.1	1.1	1.1	-0.2
Payout	0.6	-0.034	-	0.6	0.7	0.5	0.9	0.4
MLlnv2a								
Profitability	24.9	0.452	+	22.2	25.5	26.3	25.1	2.8
Growth	17.4	0.379	+	14.9	19.6	17.3	16.5	1.6
Financial soundness	16.6	0.246	+/-	22.9	16.9	12.4	14.3	-8.6
Solvency	12.3	0.571	+/-	12.8	14.1	12.2	8.8	-4.1
Efficiency	10.7	0.449	+	6.0	7.4	12.5	19.5	13.6
Industry	7.7	-0.380	N.A.	8.8	7.1	8.5	6.3	-2.5
Liquidity	4.9	0.467	+/-	3.9	4.1	6.3	5.6	1.7
Size	2.8	-0.064	-	4.2	3.0	2.4	1.5	-2.7
CAPMbeta	1.5	0.184	-	2.9	1.2	1.2	1.0	-1.9
Capitalisation	0.8	0.190	+/-	0.9	0.7	0.7	0.8	-0.1
Payout	0.4	-0.017	-	0.4	0.4	0.3	0.5	0.1
MLlnv2s								
Efficiency	36.5	-0.677	-	39.0	33.5	39.1	35.0	-4.0
Profitability	23.4	0.346	+	22.9	24.5	20.8	25.7	2.9
Growth	12.4	0.092	+	9.4	14.2	12.9	12.4	3.0
Financial soundness	8.1	0.177	+/-	7.8	8.2	7.7	9.1	1.3
Industry	5.9	0.206	N.A.	5.0	5.8	6.7	6.0	1.0
Solvency	4.9	-0.161	+/-	5.8	5.1	4.7	3.9	-1.9
Liquidity	3.3	-0.146	+/-	3.0	3.2	3.7	3.4	0.4
Size	3.1	0.090	-	4.2	3.4	2.5	2.2	-2.0
CAPMbeta	1.1	-0.118	-	1.9	1.0	0.9	0.8	-1.1
Capitalisation	0.8	-0.171	+/-	0.8	0.7	0.7	1.0	0.1
Payout	0.3	-0.026	-	0.3	0.3	0.2	0.5	0.2

Electronic copy available at: https://ssrn.com/abstract=3740270

Table 9: Importance of individual variables over time

This table reports variable importance for each of the three machine learning models (*ML_lnm2b*, *ML_lnv2a* and *ML_v2s*) for each decade from the 1980's to the 2010's. We measure the variable importance by the magnitude of *SHAP* values for each machine learning model. As an initial step we select the 10 most important variables from both the 1980 and 2010 decades for each model. Then for each model we report the 10 variables with the highest absolute amount of change (Chg = SHAP2010 minus SHAP1980). Variable definitions can be found in section A.1 of the Appendix. Section IA.2 of the Internet Appendix describes variable categories and reports summary statistics.

Variable	Description	Category	SHAP	SHAP1980	SHAP1990	SHAP2000	SHAP2010	Chg
ML_lnm2b								
profitability_an	Profitability (pbt/equity)	Profitability	6.0	2.8	6.7	5.8	9.0	6.2
rd_sale_wr	R&D/Sales	Growth	4.9	0.0	4.5	7.3	8.1	8.1
industry	Fama French 49 industries	Industry	8.9	11.1	8.7	8.3	7.4	-3.7
roe_wr	Return on Equity	Profitability	7.7	9.4	7.2	8.7	5.1	-4.3
roic_an	Return on Invested Capital	Profitability	2.9	1.1	4.0	2.6	3.4	2.4
gprof_wr	Gross Profit/Total Assets	Profitability	1.4	0.9	0.6	1.3	3.2	2.4
roce_wr	Return on Capital Employed	Profitability	2.4	0.7	1.7	4.5	2.8	2.1
aftret_invcapx_wr	After-tax Return on Invested Capital	Profitability	3.7	6.1	3.5	2.6	2.4	-3.7
CAPMbeta	CAPM beta	CAPMbeta	2.1	3.4	1.8	1.8	1.3	-2.2
chbe_an	Δ Book Equity	Growth	2.1	3.0	2.9	1.2	0.6	-2.4
ML_lnv2a								
sales2rec_an	Sales/Receivables	Efficiency	3.9	1.0	3.0	4.9	7.5	6.5
roic_an	Return on Invested Capital	Profitability	7.0	2.9	8.3	8.8	7.2	4.3
industry	Fama French 49 industries	Industry	7.7	8.8	7.1	8.5	6.3	-2.5
rd_sale_wr	R&D/Sales	Growth	3.8	0.0	3.7	5.7	6.2	6.1
rect_turn_wr	Receivables Turnover	Efficiency	1.8	0.5	0.4	2.9	4.4	3.9
roa_wr	Return on Assets	Profitability	1.8	0.4	1.5	2.5	3.3	2.9
chbe_an	Δ Book Equity	Growth	4.3	5.8	6.4	2.9	1.1	-4.7
CAPMbeta	CAPM beta	CAPMbeta	1.5	2.9	1.2	1.2	1.0	-1.9
lt_ppent_wr	Total Liabilities/PP&E	Financial soundness	4.1	5.1	7.5	1.7	0.5	-4.6
Accrual_wr	Accruals/Total Assets	Financial soundness	2.3	8.5	0.3	0.4	0.5	-8.0
ML_lnv2s								
at_turn_wr	Asset Turnover	Efficiency	9.0	7.5	7.0	9.4	13.9	6.4
opleverage_an	Operating Leverage	Efficiency	21.2	26.7	20.9	22.7	12.9	-13.9
gpm_wr	Gross Profit Margin	Profitability	2.1	1.8	1.1	2.1	4.3	2.6
ebitda2revenue_an	EBITDA Margin	Profitability	1.5	0.5	0.9	1.7	3.6	3.1
rd_sale_wr	R&D/Sales	Growth	2.5	0.0	2.7	3.8	3.3	3.3
sales2cash_an	Sales/Cash	Efficiency	2.0	0.8	1.4	2.8	3.3	2.4
npm_wr	Net Profit Margin	Profitability	3.4	5.8	2.9	2.7	2.1	-3.8
chbe_an	Δ Book Equity	Growth	2.1	2.2	3.1	1.7	1.0	-1.3
CAPMbeta	CAPM beta	CAPMbeta	1.1	1.9	1.0	0.9	0.8	-1.1
sale_ac	Sales/Turnover (Net)	Size	1.1	2.7	0.9	0.4	0.4	-2.3

Electronic copy available at: https://ssrn.com/abstract=3740270

References

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded us firms using a machine learning approach. *Journal of Accounting Research*, *58*(1), 199–235.

Bartell, R., & Janssen, C. (2017). In defense of fairness opinions. *Duff and Phelps White Paper*.

Barth, M. E., Li, K., & McClure, C. (2021). Evolution in value relevance of accounting information.

Bartram, S. M., & Grinblatt, M. (2018). Agnostic fundamental analysis works. *Journal of Financial Economics*, *128*(1), 125–147.

Bhojraj, S., & Lee, C. M. (2002). Who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, *40*(2), 407–439.

Bianchi, D., Büchner, M., & Tamoni, A. (2020). Bond risk premiums with machine learning. *The Review of Financial Studies*.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Cart. *Classification and Regression Trees; Wadsworth and Brooks/Cole: Monterey, CA, USA*.

Brennan, M. (1971). A note on dividend irrelevance and the gordon valuation model. *The Journal of Finance*, *26*(5), 1115–1121.

Brown, S., Lo, K., & Lys, T. (1999). Use of r2 in accounting research: Measuring changes in value relevance over the last four decades. *Journal of Accounting and Economics*, *28*(2), 83–115.

Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2020). *The structure of economic news* (tech. rep.). National Bureau of Economic Research.

Cao, S., Jiang, W., Wang, J. L., & Yang, B. (2021). *From man vs. machine to man+ machine: The art and ai of stock analyses* (tech. rep.). National Bureau of Economic Research.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Collins, D. W., Maydew, E. L., & Weiss, I. S. (1997). Changes in the value-relevance of earnings and book values over the past forty years. *Journal of accounting and economics*, *24*(1), 39–67.

Core, J. E., Guay, W. R., & Van Buskirk, A. (2003). Market valuations in the new economy: An investigation of what has changed. *Journal of Accounting and Economics*, *34*(1-3), 43–67.

Cragg, J. G., & Malkiel, B. G. (1968). The consensus and accuracy of some predictions of the growth of corporate earnings. *The Journal of Finance*, *23*(1), 67–84.

Damodaran, A. (2007). *Valuation approaches and metrics: A survey of the theory and evidence*. Now Publishers Inc.

Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset* (3rd edition). John Wiley & Sons.

Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, covariances, and average returns: 1929 to 1997. *The Journal of Finance*, *55*(1), 389–406.

De Franco, G., Hope, O.-K., & Larocque, S. (2015). Analysts' choice of peer companies. *Review of Accounting Studies*, *20*(1), 82–109.

Electronic copy available at: https://ssrn.com/abstract=3740270

Donelson, D. C., Jennings, R., & McInnis, J. (2011). Changes over time in the revenue-expense relation: Accounting or economics? *The Accounting Review*, *86*(3), 945–974.

Eaton, G. W., Guo, F., Liu, T., & Officer, M. S. (2021). Peer selection and valuation in mergers and acquisitions. *Available at SSRN*.

Enache, L., & Srivastava, A. (2018). Should intangible investments be reported separately or commingled with operating expenses? new evidence. *Management Science*, *64*(7), 3446–3468.

Erel, I., Stern, L. H., Tan, C., & Weisbach, M. S. (2021). Selecting directors using machine learning. *The Review of Financial Studies*, *34*(7), 3226–3264.

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252.

Francis, J., & Schipper, K. (1999). Have financial statements lost their relevance? *Journal of accounting Research*, *37*(2), 319–352.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, *38*(4), 367–378.

Geertsema, P., & Lu, H. (2020). The correlation structure of anomaly strategies. *Journal of Banking & Finance*, *119*, 105934.

Golubov, A., & Konstantinidi, T. (2019). Where is the risk in value? evidence from a market-to-book decomposition. *The Journal of Finance*, *74*(6), 3135–3186.

Gordon, M. J. (1962). The savings investment and valuation of a corporation. *The Review of Economics and Statistics*, 37–51.

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? *arXiv preprint arXiv:2207.08815*.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Hastie, T., Tibshiranie, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd edition). Springer.

Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *The Review of Financial Studies*, *23*(10), 3773–3811.

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423–1465.

Hou, K., Xue, C., & Zhang, L. (2018). Replicating Anomalies. *The Review of Financial Studies*.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 3146–3154.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767–2787.

Kim, M., & Ritter, J. R. (1999). Valuing ipos. *Journal of Financial Economics*, *53*(3), 409–437.

Electronic copy available at: https://ssrn.com/abstract=3740270

Lee, C. M., Ma, P., & Wang, C. C. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, *116*(2), 410–431.

Lev, B., & Zarowin, P. (1999). The boundaries of financial reporting and how to extend them. *Journal of Accounting research*, *37*(2), 353–385.

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal of Accounting Research*, *40*(1), 135–172.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, *2*(1), 2522–5839.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4765–4774.

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, *48*(3), 261–297.

Morgan, J. N., & Sonquist, J. A. (1963). Some results from a non-symmetrical branching process that looks for interaction effects. *Proceedings of the Social Statistics Section, American Statistical Association*, *8*, 5.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Novy-Marx, R., & Velikov, M. (2019). Comparing cost-mitigation techniques. *Financial Analysts Journal*, *75*(1), 85–102.

Rhodes-Kropf, M., Robinson, D. T., & Viswanathan, S. (2005). Valuation waves and merger activity: The empirical evidence. *Journal of Financial Economics*, *77*(3), 561–603.

Shapley, L. S. (1951). Notes on the n-person game-ii: The value of an n-person game. *Rand Corporation*.

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *The Accounting Review*, *83*(3), 823–853.

Srivastava, A. (2014). Why have measures of earnings quality changed over time? *Journal of Accounting and Economics*, *57*(2-3), 196–217.

Whitbeck, V. S., & Kisor Jr, M. (1963). A new tool in investment decision-making. *Financial Analysts Journal*, *19*(3), 55–62.

Electronic copy available at: https://ssrn.com/abstract=3740270

Appendix

A.1 Variables used in machine learning models

This table describes the construction of all the variables used in the machine learning models. Accounting information from the income statement or cash flow statement are annualised by aggregating the values from quarter $t-3$ to quarter t.

Table A.1: Variables description

Variable	Description
lnm2b	Natural logarithm of market value to book equity
lnv2a	Natural logarithm of enterprise value to total assets
lnv2s	Natural logarithm of enterprise value to sales
Accrual_wr	Accruals/Total Assets: (oancf-ib)/at
CAPMbeta	Beta on MKTRF estimated over 60 months (using CAPM)
aftret_eq_wr	After-tax Return on Common Equity: ibcom/ceq
aftret_equity_wr	After-tax Return on Stockholders' Equity: ib/seq
aftret_invcapx_wr	After-tax Return on Invested Capital: sum(ib+xint,mii)/sum(l_icapt,l_txditc,-l_mib), for postive denominator only
assetgrowth_an	Asset Growth: d_at/l_at
assets	Total book assets
assetturnover2_an	Asset Turnover: (sale/((at+l_at)/2))
assetturnover_an	Asset Turnover: sale/at
at_turn_wr	Asset Turnover: sale/at
book	Book equity as per Davis Fama French (2000)
bookequitygrowth_an	Growth in Book Equity: (be/l_be)-1
capexgrowth_an	Capex Growth: d_capx/l_capx
capital_ratio_wr	Capitalisation Ratio: dltt/(dltt+sum(ceq, pstk))
cash2assets_an	Cash/Total Assets: che/at for non-financial firms
cash_conversion_wr	Cash Conversion Cycle (Days): (invt+l_invt)/2/(cogs/365)+(rect+l_rect)/2/(sale/365)-(ap+l_ap)/2/(cogs/365)
cash_debt_wr	Cash Flow/Total Debt: ocf/lt
cash_lt_wr	Cash Balance/Total Liabilities: che/lt
cash_ratio_wr	Cash Ratio: che/lct if lct>0
cf2debt_an	Cash Flow/Total Liabilities: (ib+dp)/((lt+l_lt)/2)
cfm_wr	Cash Flow Margin: (ib+dp)/sale
chbe_an	Δ Book Equity: (be-l_be)/l_at
chca_an	Δ Non-cash Current Assets: (d_act-d_che)/l_at
chceq_an	Δ Book Equity: (ceq-l_ceq)/l_ceq
chcl_an	Δ Non-debt Current Liabilities: (d_lct-cond(missing(d_dlc),0,d_dlc))/l_at
chcurrentratio_an	Δ Current Ratio: ((act/lct)-(l_act/l_lct))/(l_act/l_lct)
chfnl_an	Δ Financial Liabilities: (cond(missing(d_dltt), 0, d_dltt) + cond(missing(d_dlc),0,d_dlc) + cond(missing(d_pstk),0,d_pstk)) / l_at
chlt_an	Δ Total Liabilities: (lt/l_lt)-1
chnccl_an	Δ Non-current Operating Liabilities: (d_lt-d_lct-cond(missing(d_dltt),0,d_dltt))/l_at
curr_debt_wr	Current Liabilities/Total Liabilities: lct/lt
curr_ratio_wr	Current Ratio: act/lct

Electronic copy available at: https://ssrn.com/abstract=3740270

Table A.1: Variables description

Variable	Description
de_ratio_wr	Total Debt/Equity : lt/sum(ceq, pstk)
debt2tang_an	Debt Capacity to Tangibility: (che+rect*0.715+invt *0.547+ppent*0.535)/at
debt_assets_wr	Total Liabilities/Total Assets: lt/at
debt_at_wr	Long-term Debt/Total Assets: (dltt+dlc)/at
debt_book_value	Book value of debt (rowtotal(dltt dlc pstk)) in millions
debt_capital_wr	Total Debt/Capital: (ap+sum(dlc,dltt))/(ap+sum(dlc,dltt)+sum(ceq,pstk))
debt_ebitda_wr	Total Debt/EBITDA: (dltt+dlc)/ebitda
debt_invcap_wr	Long-term Debt/Invested Capital: dltt/icapt if icapt > 0
deprn_an	Depreciation Rate: dp/ppent
dltt_be_wr	Long-term Debt to Equity: dltt/be if be>0
dpr_wr	Dividend Payout Ratio: dv/ibadj
ebitda2revenue_an	EBITDA Margin: ebitda/sale
efftax_wr	Effective Tax Rate: txt/pi
equity_invcap_wr	Common Equity/Invested Capital: ceq/icapt if icapt > 0
fcf_ocf_wr	Free Cash Flow/Operating Cash Flow: (ocf-capx)/ocf if ocf>0
gpm_wr	Gross Profit Margin: (sale-cogs)/sale
gprof_wr	Gross Profit/Total Assets: (sale-cogs)/at
grossmargin_an	Gross Profit Margin: (sale-cogs)/sale
grossprofit_an	Gross Profit: (sale-cogs)/at
industry	Fama French 49 industries
int_debt_wr	Interest/Long-term Debt: xint/dltt
int_totdebt_wr	Interest/Total Debt : xint/((dltt + dlc)
intcov_ratio_wr	Interest Coverage Ratio: ebit/xint
intcov_wr	After-tax Interest Coverage: (xint+ib)/xint
inventorychange_an	Δ Inventory: d_invt/(0.5*l_at+0.5*at)
inventorygrowth_an	Inventory Growth: d_invt/l_invt
investment_an	Investment: d_at/at
invt_act_wr	Inventory/Current Assets: invt/act
invturn_wr	Inventory Turnover: cogs/((invt+l_invt)/2) if (invt + l_invt)/2>0
liquid2assets_an	Liquid Assets: (che+0.75*(act-che)+0.5*(at-act))/l_at
lt_debt_wr	Long-term Debt/Total Liabilities: dltt/lt
lt_ppent_wr	Total Liabilities/PP&E: lt/ppent
marginch_an	Δ Profit Margin: (ib/sale)-(l_ib/l_sale)
npm_wr	Net Profit Margin: ib/sale
ocf_lct_wr	Operating Cash Flow/Current Liabilities: ocf/lct
opleverage_an	Operating Leverage: (cogs+xsga)/at
opmad_wr	EBIT Margin: ebit/sale
opmbd_wr	EBITDA Margin: ebitda/sale
pay_turn_wr	Payables Turnover: (cogs+d_invt)/((ap+l_ap)/2) if (ap+l_ap)/2>0
pchdeprn_an	Δ Depreciation Rate: ((dp/ppent)-(l_dp/l_ppent))/(l_dp/l_ppent)
pchgm2pchsale_an	Δ Gross Profit−Δ Sales: (((sale-cogs)-(l_sale-l_cogs))/(l_sale-l_cogs))-((sale-l_sale)/l_sale)
pchquickratio_an	Δ Quick Ratio: ((act-invt)/lct-(l_act-l_invt)/l_lct)/((l_act-l_invt)/l_lct)
pchsale2pchinvt_an	Δ Sales−Δ Inventory: ((sale-l_sale)/l_sale)-((invt-l_invt)/l_invt)
pchsale2pchrect_an	Δ Sales−Δ Receivables: ((sale-l_sale)/l_sale)-((rect-l_rect)/l_rect)
pchsale2pchxsga_an	Δ Sales−Δ SGA: ((sale- l_sale)/l_sale)-((xsga-l_xsga)/l_xsga)
pchsales2inv_an	Δ Sales/Inventory: ((sale/invt)-(l_sale/l_invt))/(l_sale/l_invt)
pretret_earnat_wr	Pre-tax Return on Total Earning Assets: ebitda/earnat; Total Earning Assets: ppent+act
pretret_noa_wr	Pre-tax Return on Net Operating Assets: ebit/noa; Net operating assets: ppent+act-lct
profit_lct_wr	EBITDA/Current liabilities: ebitda/lct

Electronic copy available at: https://ssrn.com/abstract=3740270

Table A.1: Variables description

Variable	Description
profitability_an	Profitability: (sale-cogs-xint-xsga)/be
ptpm_wr	Pre-tax Profit Margin: (ebit-xint+spi+nopi)/sale
quick_ratio_wr	Quick Ratio: (act-invt)/lct if lct>0
rd_sale_wr	R&D/Sales: max(xrd/sale,0), missing xrd replaced with zero
rect_act_wr	Receivables/Current Assets: rect/act
rect_turn_wr	Receivables Turnover: sale/((rect+l_rect)/2)
roa_wr	Return on Assets: ebitda/at
roce_wr	Return on Capital Employed: ebit/(dltt + pstk + dlc + ceq)
roe_wr	Return on Equity: ib/be if be>0
roic_an	Return on Invested Capital: (ebit-nopi)/(ceq+lt-che)
sale_ac	Sales/Turnover (Net)
sales2cash_an	Sales/Cash: sale/che
sales2inv_an	Sales/Inventory: sale/invt
sales2rec_an	Sales/Receivables: sale/rect
salesgrowth_an	Sales Growth: (sale/l_sale)-1
short_debt_wr	Short-Term Debt/Total Debt: dlc/(dltt+dlc)
totdebt_invcap_wr	Total Debt/Invested Capital: (dltt+dlc)/icapt if icapt > 0

A.2 Gradient boosting machines (GBMs)

GBMs are a type of tree-based prediction method. Tree-based prediction methods proceed by finding a sequence of rules that allocate training observations to leaves in such a way that the leaf average target value is a good predictor.[19] In this sub-section we use a simplified example to demonstrate how GBMs are trained on training data to make out-of-sample predictions for test data.

The training data set in our example contains five firms (labelled A to E). The two input variables are ROE and AssetGrowth. The target variable in the example is the log of the market-to-book multiple (*lnm2b*). For ease of illustration the test data set contains only one firm (labelled X). Panel A of Table A.2 summarises the training and test data sets.

A GBM consists of a set of sequentially constructed tree predictors. In our example the GBM consists of two trees, each with a single decision node and two terminal leaves. The decision node consists of an inequality between an input variable and a threshold, both selected during the training process to minimise the squared errors of tree predictions. Observations are allocated to the terminal ("leaf") nodes of a tree by following the decision rules in the

[19]The decision-tree based non-parametric estimation approach was introduced by Morgan and Sonquist (1963) and later popularised by (Breiman, Friedman, Olshen et al., 1984) as the Classification and Regression Trees approach.

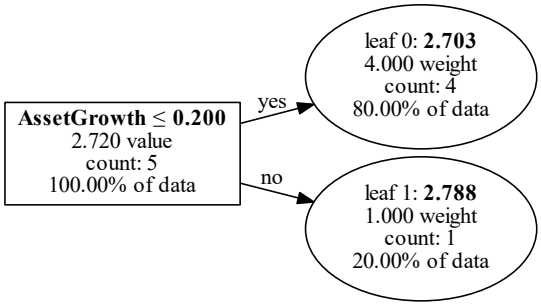
Electronic copy available at: https://ssrn.com/abstract=3740270

tree. The prediction from a tree is the average of the target values of training observations allocated to a particular leaf node. The resulting trees are presented in Figure A.1. Panel B of Table A.2 presents the leaf allocations of all the observations to one of the two possible leaves (labelled 0 and 1 respectively) for both Tree 1 and Tree 2.

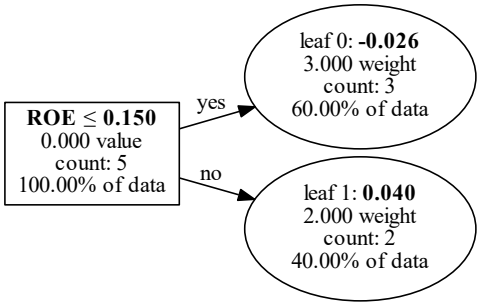
Figure A.1: Trees in GBMs: a simplified example

These figures depict *tree 1* and *tree 2* in GBMs from the simplified example in Table A.2. A rectangular frame denotes a decision node which contains the decision rule, the total count and percentage of training data subject to that decision rule. An oval frame denotes a leaf containing the prediction (in a bold font) and observation counts. The prediction in panel A consists of the base sample average of 2.72 plus the tree 1 prediction scaled by the learning rate of 0.1. The prediction in panel B consists of the tree 2 prediction scaled by the learning rate. The GBM prediction can be obtained directly by adding the predictions of the trees in panels A and B.

A: *Tree 1*



B: *Tree 2*



Before adding any trees, in iteration 0 the GBM prediction is set to equal to the sample average of the target value in the training dataset (2.72). Every following iteration corresponds to the addition of a further tree to the GBM. At each iteration the GBM prediction is equal to the sum of the initial iteration 0 prediction (the training data target value average) and the predictions from all the trees added to the GBM at that point, each scaled by a fraction called the *learning rate* (0.1 in our example).

Each new tree is trained on the *residuals* of the previous iteration GBM prediction for the training dataset (see Panel E), calculated as the difference between the actual target value and the GBM predicted target value.[20] Panel C of Table A.2 presents the predictions for iteration 0 (the training data target value sample average of 2.72), iteration 1 (corresponding

[20]In general GBMs train each new tree on the gradient of the current GBM model (hence the name *gradient* boosting). When the target variable is continuous and least squares loss function is used (as in this paper) the gradient is equal to the residuals.

Electronic copy available at: https://ssrn.com/abstract=3740270

to the prediction of tree 1 times the learning rate of 0.1) and iteration 2 (corresponding to the prediction of tree 2, again scaled by the learning rate of 0.1). The GBM prediction at each iteration is set out in Panel D. Consider the GBM prediction for firm A at iteration 2 of 2.743. This prediction is calculated as the sum of the training data sample average of 2.72 and the two tree predictions for firm A (each already scaled by the learning rate of 0.1) as per Panel C: $2.72 + (-0.017) + 0.040 = 2.742 \approx 2.743$ (difference due to rounding).

We now consider calculations at the individual tree level. In iteration 1, each observation is allocated to a single leaf based on the optimal decision rule. Firm A is allocated to leaf 0 because its AssetGrowth is 18% ($\leq$ 20%) (Panel A of Figure A.1). The prediction for Firm A in this iteration is the average of residuals of firms on leaf 0 times the learning rate ($\frac{0.88-0.12-0.52-0.92}{4} \times 0.1 = -0.017$; refer to Panels B, E and C of Table A.2). The GBM prediction for Firm A in iteration 1 is 2.703, or the GBM prediction from iteration 0 plus the tree prediction in iteration 1 scaled by the learning rate ($2.72 + (-0.017) = 2.703$).

Finally, we predict the target value of the out-of-sample Firm X. The algorithm starts from the training sample average target value of 2.72 as the initial GBM prediction for iteration 0. Then, it adjusts this initial GBM prediction by adding the tree prediction for the Firm X observation in each further iteration. Firm X in our example is predicted to have a target value of 2.743 after two iterations (calculated as $2.72 - 0.017 + 0.04 = 2.743$ from Panels C and D of Table A.2).

Electronic copy available at: https://ssrn.com/abstract=3740270

Table A.2: Using GBMs to generate predictions in test data: a simplified example

These tables summarise inputs, intermediate steps and outputs from a simplified example using GBMs to make predictions. Panel A summarises the training and test data used in the GBMs. *Train* equals to one for training data and 0 otherwise. Panel B reports the leaf allocation for all observations. Panels C, D and E tabulate the predictions, GBM predictions and residuals in each iteration, respectively.

A: Training and test data

Firm	Data	lnm2b	ROE	AssetGrowth
A	Train	3.6	16%	18%
B	Train	3.4	13%	22%
C	Train	2.6	16%	10%
D	Train	2.2	4%	12%
E	Train	1.8	14%	18%
X	Test	2.6	25%	5%

B: Leaf allocation in each tree

Firm	Data	Tree 1	Tree 2
A	Train	0	1
B	Train	1	0
C	Train	0	1
D	Train	0	0
E	Train	0	0
X	Test	0	1

C: Tree predictions

Firm	Data	Iteration 0	Iteration 1	Iteration 2
A	Train	2.72	-0.017	0.040
B	Train	2.72	0.068	-0.026
C	Train	2.72	-0.017	0.040
D	Train	2.72	-0.017	-0.026
E	Train	2.72	-0.017	-0.026
X	Test	2.72	-0.017	0.040

D: GBM predictions

Firm	Data	Iteration 0	Iteration 1	Iteration 2
A	Train	2.72	2.703	2.743
B	Train	2.72	2.788	2.762
C	Train	2.72	2.703	2.743
D	Train	2.72	2.703	2.677
E	Train	2.72	2.703	2.677
X	Test	2.72	2.703	2.743

E: Residuals

Firm	Data	Iteration 0	Iteration 1
A	Train	0.88	0.897
B	Train	0.68	0.612
C	Train	-0.12	-0.103
D	Train	-0.52	-0.503
E	Train	-0.92	-0.903

A.3 Explainable GBM predictions with peer weights

We show that any GBM prediction can be expressed as a weighted average of training target values.[21] In this subsection we set out the calculations required to obtain the *peer weights* for out-of-sample test observations. In the following discussion *firms* and *observations* are equivalent, since each GBM is trained and applied in a particular month.

[21] See https://github.com/pgeertsema/peerweights for Python code to implement this for LightGBM predictions.

Electronic copy available at: https://ssrn.com/abstract=3740270

We first calculate the in-sample peer weights of training firms that generate the equivalent GBM predictions in each iteration.

The initial prediction is the training sample average target value of all training observations; or, all firms in the training sample carry an equal weight of $\frac{1}{N}$ in iteration 0. The tree prediction weight matrix ($\boldsymbol{P}_0$) and associated *tree predictions* ($\boldsymbol{p}_0$) in iteration $i = 0$ are:

$$\boldsymbol{P}_0 := \frac{1}{N}\mathbf{1}_{N\times N} \tag{11}$$

$$\boldsymbol{p}_0 := \boldsymbol{v}^T\boldsymbol{P}_0 \tag{12}$$

where $\boldsymbol{v}$ is a $N \times 1$ column vector of the target values of the N training firms. Equation (12) initialises in-sample tree predictions to the mean of the training firm target values.

The GBM weight matrix ($\boldsymbol{G}_0$) and associated the *GBM predictions* ($\boldsymbol{g}_0$) in iteration $i = 0$ are given by:

$$\boldsymbol{G}_0 := \boldsymbol{P}_0 \tag{13}$$

$$\boldsymbol{g}_0 := \boldsymbol{v}^T\boldsymbol{G}_0 \tag{14}$$

We now consider the iterations $i \in [1, ..., M]$ where M is the number of trees in the GBM model. Steps 1 to 3 below are repeated for each $i \in [1, ..., M]$.

Step 1: The leaf weight matrix $\boldsymbol{W}_i$ is an $N \times N$ matrix given by (15)

$$\boldsymbol{W}_i = \boldsymbol{D}_i \oslash (\mathbf{1}_{N\times N}\boldsymbol{D}_i) \tag{15}$$

where each element $d_{i,f,c}$ in $\boldsymbol{D}_i$ takes the value one if firm f and firm c are allocated to the same leaf in iteration i, otherwise zero. $\mathbf{1}_{\boldsymbol{N\times N}}$ is an $N \times N$ matrix where each element is 1. The $\oslash$ operator denotes element-by-element division. In $\boldsymbol{W}_i$, if firm c and firm f are allocated to the same leaf, $w_{i,f,c}$ takes the value of one over the total number of firms

Electronic copy available at: https://ssrn.com/abstract=3740270

allocated to that leaf. If firm c and firm f are allocated to different leaves, $w_{i,f,c}$ takes the value of zero.

Step 2: Use the residual weight matrix from the previous iteration ($\boldsymbol{I} - \boldsymbol{G}_{i-1}$), the leaf weight matrix $\boldsymbol{W}_i$ at the current iteration and learning rate λ (set to the default value of 0.1 in our study) to calculate the tree prediction weight matrix ($\boldsymbol{P}_i$) and the associated tree predictions scaled by the learning rate λ ($\boldsymbol{p}_i$)

$$\boldsymbol{P}_i := \lambda \boldsymbol{W}_i(\boldsymbol{I} - \boldsymbol{G}_{i-1}) \tag{16}$$

$$\boldsymbol{p}_i := \boldsymbol{v}^T \boldsymbol{P}_i^T \boldsymbol{W}_i \tag{17}$$

Each element $p_{i,f,c}$ in $\boldsymbol{P}_i$ is the weight on training firm c used to calculate the tree prediction of firm f in iteration i. The initial residual matrix ($\boldsymbol{I} - \boldsymbol{G}_0$) contains identical and positive diagonal elements ($\frac{N-1}{N}$) and identical and negative off-diagonal elements ($-\frac{1}{N}$).[22]

Step 3: Add the tree prediction weight matrix at the current iteration ($\boldsymbol{P_i}$) to the GBM prediction weight matrix in the previous iteration ($\boldsymbol{G_{i-1}}$) to generate the GBM prediction weight matrix ($\boldsymbol{G_i}$) used for calculating the GBM predictions ($\boldsymbol{g_i}$):

$$\boldsymbol{G}_i := \boldsymbol{G}_{i-1} + \boldsymbol{P}_i \tag{18}$$

$$\boldsymbol{g}_i := \boldsymbol{v}^T \boldsymbol{G}_i \tag{19}$$

Each element $g_{i,f,c}$ of $\boldsymbol{G_i}$, is the weight on the component firm c used to calculate the GBM prediction of firm f in iteration i.

Finally, we express the out-of-sample GBM predictions $\boldsymbol{k}$ for S test firms (an $S \times 1$ column vector) as weighted averages of training firm valuation multiples. $\boldsymbol{K}$ is a $N \times S$ matrix containing the column weights used to generate $\boldsymbol{k}$. $\boldsymbol{k}$ and $\boldsymbol{K}$ are defined as

$$\boldsymbol{K} := \sum_{i=0}^{M} (\boldsymbol{P}_i^T (\boldsymbol{L}_i \oslash (\boldsymbol{1}_{N \times N} \boldsymbol{L}_i)) \tag{20}$$

[22] Recall that $\boldsymbol{I}$ is the identity matrix, whose off-diagonal elements are zero and diagonal elements are 1.

Electronic copy available at: https://ssrn.com/abstract=3740270

$$\boldsymbol{k} := \boldsymbol{v}^T \boldsymbol{K} \tag{21}$$

where $\boldsymbol{L_i}$ is a $N \times S$ indicator matrix in iteration i recording the leaf location of test firms, playing the same role as D_i in equation (15). Each element $l_{i,s,f}$ ($f \in [1 \cdots N]$ and $s \in [1 \cdots S]$) in $\boldsymbol{L_i}$ takes the value of one if the test firm s is allocated to the same leaf as the training firm f in iteration i in the tree prediction weight matrix $\boldsymbol{P_i}$, and zero otherwise.

In equation (20), the final peer weights $\boldsymbol{K}$ are linear combinations of the corresponding elements in $\boldsymbol{P_i}$ accumulated from iteration 0 to iteration M.

We validate the algorithm described above in two ways.

First, we use the simplified example in Section A.2. In this simplified example the predicted *lnm2b* for the test firm X is equivalent to the weighted average of the peer firm targets. The weights on firms A, B, C, D and E are 0.235, 0.162, 0.235, 0.185 and 0.185, respectively. Applying these weights to *lnm2b* of training firms in Panel A of Table A.2, we generate a predicted *lnm2b* of 2.743. This is identical to the GBM output after iteration 2 for firm X in Panel D of Table A.2.

Second, we validate our algorithm using the December 2019 data. We confirm that a weighted average of peers' valuation multiples, as calculated above, equals the GBM predicted valuation multiples.

Electronic copy available at: https://ssrn.com/abstract=3740270

Internet Appendix – Relative Valuation with Machine Learning

Electronic copy available at: https://ssrn.com/abstract=3740270

IA.1 Traditional valuation models

This table describes the construction of variables used in the traditional valuation model based on Rhodes-Kropf, Robinson and Viswanathan (2005).

Table IA.1: Rhodes-Kropf, Robinson and Viswanathan (2005)

Variable	Description
lnm	Natural logarithm of market value of equity from CRSP (ln(crsp_market_equity))
lnb	Natural logarithm of book equity(ln(be_ac))
lnni	Natural logarithm of absolute value of the last four-quarter net income (ln(abs(ni_ac)))
negni	An indicator variable for loss (ni_ac < 0 & ni_ac != .)
LEV1	Leverage ((dltt+dlc)/at) as in Golubov and Konstantinidi (2019)
LEV2	Book leverage (1-book/at) as in Rhodes-Kropf, Robinson and Viswanathan (2005)
LEV3	Market leverage (1 - m2b * book / (m2b * book + debt_book_value)) as in Rhodes-Kropf, Robinson and Viswanathan (2005)

Electronic copy available at: https://ssrn.com/abstract=3740270

This table describes the construction of variables used in the traditional valuation model based on Bartram and Grinblatt (2018).

Table IA.2: Bartram and Grinblatt (2018)

Variable	Description
at	Assets - Total
seq	Stockholders Equity > Parent > Index Fundamental > Quarterly
icapt	Invested Capital - Total - Quarterly
teq	Stockholders Equity - Total
pstkr	Preferred/Preference Stock - Redeemable
ppent	Property Plant and Equipment - Total (Net)
ceq	Common/Ordinary Equity - Total
pstk	Preferred/Preference Stock (Capital) - Total
dltt	Long-Term Debt - Total
ao	Assets - Other - Total
lt	Liabilities - Total
lo	Liabilities - Other
che	Cash and Short-Term Investments
aco	Current Assets - Other - Total
lco	Current Liabilities - Other - Total
ap	Account Payable/Creditors - Trade
dv	cash dividend
dvp	Dividends - Preferred/Preference
sale	Sales/Turnover (Net)
ib	Income Before Extraordinary Items
ni	Net Income (Loss)
xido	Extraordinary Items and Discontinued Operations
ibadj	Income Before Extraordinary Items - Adjusted for Common Stock Equivalents
ibcom	Income Before Extraordinary Items - Available for Common
pi	Pretax Income
txt	Income Taxes - Total
nopi	Non-Operating Income (Expense) - Total
do	Discontinued Operations

Electronic copy available at: https://ssrn.com/abstract=3740270

This table describes the construction of variables used in the traditional valuation model based on Bhojraj and Lee (2002).

Table IA.3: Bhojraj and Lee (2002)

Variable	Description
m2b	Market value to book equity
v2a	Enterprise value to total assets
v2s	Firm value (market value of equity + total debt) to sales (sale) multiple
indm2b	Cross-sectional industry harmonic mean of market to book multiple
indv2a	Cross-sectional industry harmonic mean of firm value to total assets multiple
indv2s	Cross-sectional industry harmonic mean of firm value to sales multiple
adjopmad	Industry adjusted opearting profit margain (actual - median); opmad_wr-opmad_wm
negadjopmad	adjopmad * an indicator variable for negative adjopmad
adjsalesgrowth	Industry adjusted sales growth (actual - median)
d2e	Total debt (dltt + dlc) to book equity
pretret_noa_wr	Pre-tax Return on Net Operating Assets: ebit/noa; Net operating assets: ppent+act-lct
roe_wr	Return on Equity: ib/be if be>0
rd_sale_wr	R&D/Sales: max(xrd/sale,0), missing xrd replaced with zero

Electronic copy available at: https://ssrn.com/abstract=3740270

IA.2 Summary statistics of variables used in machine learning models

This table presents summary statistics of all variables used in machine learning models by category.

Table IA.4: Summary statistics of variables used in machine learning models

Category	Variable	N	Mean	p1	p10	p50	p90	p99
CAPMbeta	CAPMbeta	1,325,385	1.06	-0.20	0.30	0.99	3.19	3.19
Capitalisation	capital_ratio_wr	1,793,810	0.29	0.00	0.00	0.26	0.92	0.92
Capitalisation	debt_invcap_wr	1,793,926	0.29	0.00	0.00	0.25	0.93	0.93
Capitalisation	equity_invcap_wr	1,804,413	0.71	0.09	0.35	0.74	1.00	1.00
Capitalisation	totdebt_invcap_wr	1,701,166	0.45	0.00	0.00	0.36	2.48	2.48
Efficiency	assetturnover2_an	1,718,671	1.09	0.05	0.10	0.94	4.31	4.31
Efficiency	assetturnover_an	1,811,786	1.11	0.04	0.09	0.89	4.20	4.20
Efficiency	at_turn_wr	1,811,786	1.11	0.04	0.09	0.89	4.20	4.20
Efficiency	invturn_wr	1,378,546	27.01	0.30	1.78	5.16	262.62	262.62
Efficiency	opleverage_an	1,811,786	1.00	0.02	0.07	0.78	4.07	4.07
Efficiency	pay_turn_wr	1,621,084	17.38	0.00	0.12	8.74	76.87	76.87
Efficiency	pchgm2pchsale_an	1,786,172	-0.67	-4.29	-0.25	-0.00	2.92	2.92
Efficiency	pchsale2pchinvt_an	1,361,650	-0.45	-4.95	-0.49	0.01	1.47	1.47
Efficiency	pchsale2pchrect_an	1,650,893	-1.55	-2.53	-0.39	-0.00	1.44	1.44
Efficiency	pchsale2pchxsga_an	1,512,508	0.42	-1.40	-0.21	-0.00	1.68	1.68
Efficiency	pchsales2inv_an	1,348,862	0.56	-0.82	-0.32	0.01	4.56	4.56
Efficiency	rect_turn_wr	1,655,563	16.77	0.07	0.18	6.10	178.23	178.23
Efficiency	sales2cash_an	1,791,692	168.90	0.25	1.11	10.29	1644.15	1644.15
Efficiency	sales2inv_an	1,441,369	52.75	0.69	3.12	8.11	436.99	436.99
Efficiency	sales2rec_an	1,752,533	33.47	0.06	0.17	5.82	182.86	182.86
Financial soundness	Accrual_wr	1,811,786	0.03	-0.23	-0.08	0.01	0.45	0.45
Financial soundness	cash_debt_wr	1,811,467	0.22	-1.15	-0.04	0.04	1.57	1.57
Financial soundness	cash_lt_wr	1,802,806	0.61	0.00	0.01	0.11	7.13	7.13
Financial soundness	curr_debt_wr	1,482,851	0.55	0.08	0.20	0.53	1.00	1.00
Financial soundness	debt_ebitda_wr	1,574,290	-8.4e+11	-22.17	0.00	1.68	35.49	35.49
Financial soundness	dltt_be_wr	1,794,070	0.81	0.00	0.00	0.33	7.66	7.66
Financial soundness	fcf_ocf_wr	1,071,032	-157.43	-12.38	-0.88	0.58	1.00	1.00
Financial soundness	int_debt_wr	1,529,229	0.68	0.00	0.00	0.08	5.07	5.07
Financial soundness	int_totdebt_wr	1,516,415	0.26	0.00	0.00	0.06	0.89	0.89
Financial soundness	invt_act_wr	1,450,025	0.28	0.00	0.00	0.26	0.87	0.87
Financial soundness	lt_debt_wr	1,794,003	0.29	0.00	0.00	0.26	0.87	0.87
Financial soundness	lt_ppent_wr	1,779,690	20.30	0.30	0.74	2.26	215.65	215.65
Financial soundness	ocf_lct_wr	1,482,849	0.33	-1.92	-0.11	0.19	2.71	2.71
Financial soundness	profit_lct_wr	1,482,849	0.99	-2.20	-0.01	0.51	3.55	3.55
Financial soundness	rect_act_wr	1,461,920	0.37	0.00	0.10	0.36	0.88	0.88
Financial soundness	short_debt_wr	1,516,415	0.28	0.00	0.00	0.15	1.00	1.00
Growth	assetgrowth_an	1,718,671	3.08	-0.39	-0.10	0.08	2.95	2.95
Growth	bookequitygrowth_an	1,718,686	1.14	-1.60	-0.18	0.09	5.26	5.26
Growth	capexgrowth_an	1,382,603	163.16	-1.00	-0.60	0.10	17.53	17.53
Growth	chbe_an	1,718,508	1.85	-0.33	-0.07	0.03	2.14	2.14

Electronic copy available at: https://ssrn.com/abstract=3740270

Table IA.4: Summary statistics of variables used in machine learning models

Category	Variable	N	Mean	p1	p10	p50	p90	p99
Growth	chca_an	1,400,967	0.54	-0.23	-0.06	0.02	0.84	0.84
Growth	chceq_an	1,709,220	0.99	-1.79	-0.18	0.08	5.35	5.35
Growth	chcl_an	1,409,549	0.51	-0.15	-0.04	0.01	0.51	0.51
Growth	chfnl_an	1,404,165	0.55	-0.41	-0.08	0.00	1.05	1.05
Growth	chlt_an	1,717,885	0.82	-0.55	-0.17	0.07	4.45	4.45
Growth	chnccl_an	1,409,410	0.21	-0.09	-0.01	0.00	0.20	0.20
Growth	deprn_an	1,780,029	0.36	0.00	0.01	0.15	2.08	2.08
Growth	inventorychange_an	1,654,510	0.01	-0.15	-0.02	0.00	0.26	0.26
Growth	inventorygrowth_an	1,362,107	0.74	-0.99	-0.29	0.07	5.76	5.76
Growth	investment_an	1,718,671	0.08	-0.64	-0.11	0.07	0.75	0.75
Growth	pchdeprn_an	1,510,550	0.26	-1.00	-0.35	0.02	3.47	3.47
Growth	rd_sale_wr	1,811,786	0.04	0.00	0.00	0.00	0.69	0.69
Growth	salesgrowth_an	1,786,178	0.83	-0.66	-0.14	0.09	3.35	3.35
Industry	industry	1,792,062	31.11	2.00	11.00	35.00	48.00	48.00
Liquidity	cash2assets_an	1,401,669	0.15	0.00	0.01	0.07	0.79	0.79
Liquidity	cash_conversion_wr	1,589,855	-1626.02	-3.9e+04	-1096.70	61.01	754.06	754.06
Liquidity	cash_ratio_wr	1,479,424	1.10	0.00	0.03	0.32	10.55	10.55
Liquidity	chcurrentratio_an	1,403,692	0.13	-0.72	-0.34	-0.01	2.78	2.78
Liquidity	curr_ratio_wr	1,479,469	2.77	0.43	0.95	2.02	13.43	13.43
Liquidity	liquid2assets_an	1,405,219	2.60	0.39	0.54	0.71	3.09	3.09
Liquidity	pchquickratio_an	1,370,712	0.16	-0.75	-0.38	-0.01	3.36	3.36
Liquidity	quick_ratio_wr	1,475,700	2.09	0.18	0.56	1.35	12.22	12.22
Payout	dpr_wr	1,401,994	1.2e+10	0.00	0.00	0.00	3.03	3.03
Profitability	aftret_eq_wr	1,804,542	0.02	-1.65	-0.19	0.09	0.66	0.66
Profitability	aftret_equity_wr	1,811,533	0.03	-1.51	-0.18	0.09	0.63	0.63
Profitability	aftret_invcapx_wr	1,714,245	0.24	-0.71	-0.08	0.09	0.71	0.71
Profitability	cfm_wr	1,811,786	0.05	-1.17	-0.04	0.08	0.51	0.51
Profitability	ebitda2revenue_an	1,811,786	0.12	-0.89	0.00	0.12	0.65	0.65
Profitability	efftax_wr	1,416,768	1.8e+10	-1.01	0.11	0.36	1.35	1.35
Profitability	gpm_wr	1,811,786	0.35	-0.43	0.11	0.34	0.91	0.91
Profitability	gprof_wr	1,811,786	0.33	-0.16	0.04	0.27	1.17	1.17
Profitability	grossmargin_an	1,811,786	0.36	-0.39	0.12	0.35	0.91	0.91
Profitability	grossprofit_an	1,811,786	0.34	-0.12	0.04	0.27	1.18	1.18
Profitability	marginch_an	1,786,178	0.25	-0.76	-0.09	0.00	1.21	1.21
Profitability	npm_wr	1,811,786	-0.00	-1.37	-0.10	0.04	0.35	0.35
Profitability	opmad_wr	1,811,786	0.07	-1.20	-0.05	0.09	0.55	0.55
Profitability	opmbd_wr	1,811,786	0.12	-0.89	0.00	0.12	0.65	0.65
Profitability	pretret_earnat_wr	1,476,641	0.11	-0.55	-0.07	0.10	0.56	0.56
Profitability	pretret_noa_wr	1,474,245	8.6e+08	-1.28	-0.12	0.14	1.52	1.52
Profitability	profitability_an	1,811,786	0.23	-0.93	-0.06	0.21	1.31	1.31
Profitability	ptpm_wr	1,811,786	0.03	-1.43	-0.11	0.07	0.48	0.48
Profitability	roa_wr	1,811,786	0.10	-0.30	0.00	0.09	0.39	0.39
Profitability	roce_wr	1,695,818	0.11	-0.57	-0.06	0.11	0.53	0.53
Profitability	roe_wr	1,811,786	0.03	-1.52	-0.17	0.09	0.60	0.60
Profitability	roic_an	1,796,195	0.23	-1.04	-0.07	0.07	0.60	0.60
Size	assets	1,812,266	5260.28	6.79	28.48	333.26	77207.00	77207.00
Size	book	1,812,266	1188.75	3.22	13.69	123.29	18675.00	18675.00
Size	debt_book_value	1,812,266	1490.82	0.00	0.00	47.79	18617.00	18617.00
Size	sale_ac	1,812,266	2067.93	5.80	23.68	220.62	32957.90	32957.90
Solvency	cf2debt_an	1,717,912	0.16	-1.66	-0.06	0.12	1.71	1.71
Solvency	de_ratio_wr	1,802,670	3.11	0.08	0.27	1.25	20.74	20.74

Electronic copy available at: https://ssrn.com/abstract=3740270

Table IA.4: Summary statistics of variables used in machine learning models

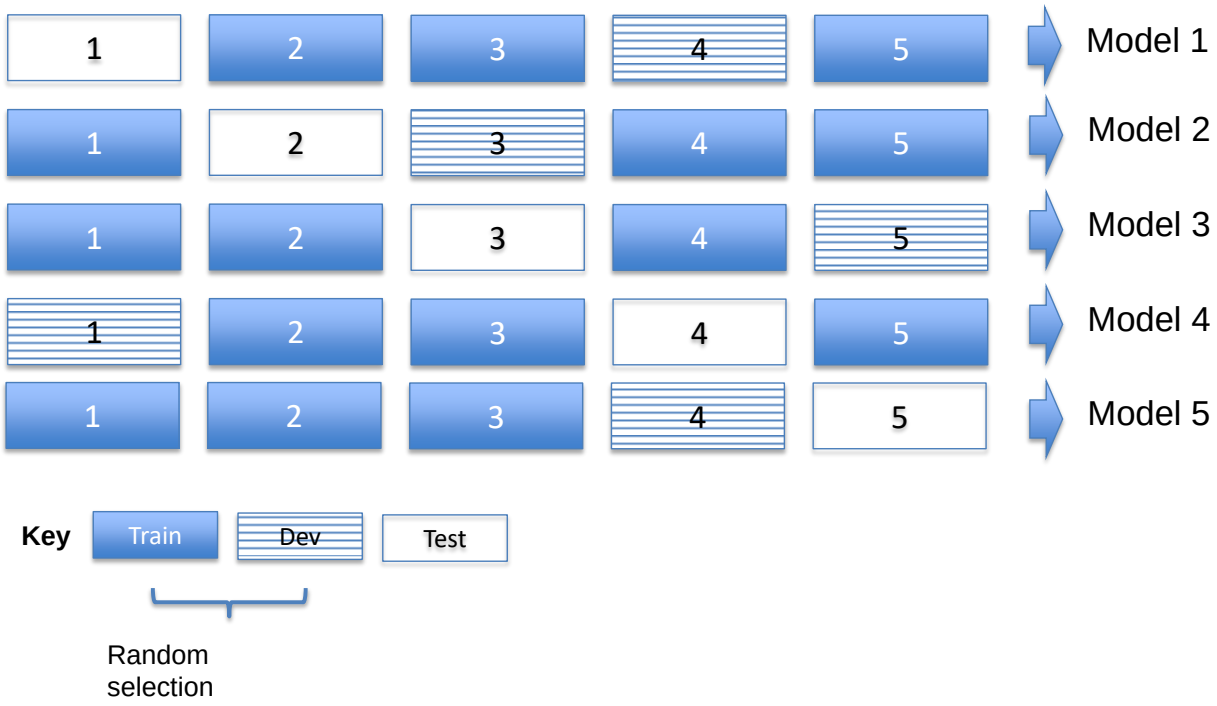
Category	Variable	N	Mean	p1	p10	p50	p90	p99
Solvency	debt2tang_an	1,698,813	0.52	0.11	0.32	0.53	0.88	0.88
Solvency	debt_assets_wr	1,811,428	0.55	0.08	0.21	0.55	0.95	0.95
Solvency	debt_at_wr	1,701,277	0.22	0.00	0.00	0.19	0.74	0.74
Solvency	debt_capital_wr	1,770,659	0.46	0.01	0.08	0.45	0.96	0.96
Solvency	intcov_ratio_wr	1,325,080	42.93	-138.90	-1.81	3.68	717.17	717.17
Solvency	intcov_wr	1,325,080	23.12	-166.05	-2.58	2.77	517.00	517.00
Target variable	lnm2b	1,812,266	0.53	-1.51	-0.45	0.48	2.87	2.87
Target variable	lnv2a	1,812,266	-0.02	-2.16	-1.18	-0.03	2.11	2.11
Target variable	lnv2s	1,812,266	0.38	-1.96	-0.92	0.35	2.92	2.92

Electronic copy available at: https://ssrn.com/abstract=3740270

IA.3 Data splitting approach – training, validation and test datasets

Figure IA.1: Training, validation and test data split and rotation

This figure illustrates the split of data into training, validation and test datasets in *each* of the 480 monthly cross-sections (January 1980 to December 2019) for machine learning models. Firms in the full sample (as described in Table 1) are assigned to five blocks using a pre-selected random seed. To generate predicated target values for all observations, each of the five blocks is used as the test data once. Then, we randomly choose a validation block from the remaining blocks for machine learning models. The other three blocks are the training blocks for the machine learning model. In estimating non-machine-learning models, we use four blocks of data as in-sample data to make predictions for observations in the out-of-sample block. Similarly, each of the five blocks is used to make out-of-sample predictions once. Five models are trained (fitted in sample) for each target (dependent) variable using a machine-learning (non-machine learning) model in each month. A total of 2,400 models are trained (estimated) for each target (dependent) variable.



Electronic copy available at: https://ssrn.com/abstract=3740270

IA.4 Translating model prediction to equity value

If the target variable is the log of the market-to-book ratio, the value of equity is estimated as

$$\widehat{M} = e^{\widehat{lnm2b}} \times B \tag{22}$$

where $\widehat{lnm2b}$ is the predicted log of the market-to-book ratio and B is the book value of equity (constructed following the approach in Davis, Fama and French, 2000).

If the target variable is the log of the enterprise-value-to-assets multiple, the value of equity is estimated as

$$\widehat{M} = \left(e^{\widehat{lnv2a}} \times A\right) - D \tag{23}$$

where $\widehat{lnv2a}$ is the predicted log of the enterprise-value-to-assets multiple, A is total assets (atq) and D is the book value of total debt (the sum of $dltt$, dlc and $pstk$, each zero if missing).

If the target variable is the log of the enterprise-value-to-sales multiple, the value of equity is estimated as

$$\widehat{M} = \left(e^{\widehat{lnv2s}} \times S\right) - D \tag{24}$$

where $\widehat{lnv2s}$ is the predicted log of the enterprise-value-to-sales multiple and S is annualised sales ($\sum_{i=0}^{3} saleq_{t-i}$).

In addition, the *RRV* model predicts the log of equity value ($\widehat{lnm}$), so we obtain $\widehat{M}$ as $e^{\widehat{lnm}}$. The outputs from the *BL* models and *IND* models are valuation multiples, so the value of equity is calculated from the estimated valuation multiples as follows:

$$\widehat{M} = \widehat{m2b} \times B \tag{25}$$

Electronic copy available at: https://ssrn.com/abstract=3740270

$$\widehat{M} = \widehat{v2a} \times A - D \tag{26}$$

$$\widehat{M} = \widehat{v2s} \times S - D \tag{27}$$

Electronic copy available at: https://ssrn.com/abstract=3740270

IA.5 Valuation error benchmarking (equal number of observations across models)

GBMs can generate more predictions than traditional models because it does not require all input variables to have non-missing values to generate a prediction for an observation. In the main text, we present valuation accuracy based on all the predictions from a model, so the sample size varies between models. In this section, we report valuation accuracy benchmarking results when the sample is the same for all models by only considering observations with no missing predictions from any model (except for the HP models, because HP data are only available from 1989 onwards).

Electronic copy available at: https://ssrn.com/abstract=3740270

Electronic copy available at: https://ssrn.com/abstract=3740270

Table IA.5: Model performance comparison

This table reports summary statistics of valuation errors from different models. This table compares valuation accuracy for observations with valid predictions from all models (except HP models). Refer to Table 2 for a description of the methodology.

	N	med($\lvert\varepsilon\rvert$)	avg($\lvert\varepsilon\rvert$)	med(ε)	avg(ε)	sd(ε)	rmse($\hat{y}$)	rmse($\hat{M}$)	$r^2_{oos}(\hat{y})$	$r^2_{oos}(\hat{M})$	$\rho_{M,\hat{M}}$
ML_lnm2b	1,262,323	0.31	0.50	-0.00	0.19	1.13	0.58	7,683	0.54	0.77	0.89
ML_lnv2a	1,262,323	0.31	0.53	0.01	0.24	1.35	0.44	7,355	0.62	0.79	0.90
ML_lnv2s	1,262,323	0.32	0.56	0.01	0.25	1.40	0.45	7,235	0.80	0.79	0.91
BL_m2b	1,262,323	0.52	1.15	0.21	0.73	3.89	6.33	12,572	0.12	0.37	0.79
BL_v2a	1,262,323	0.67	1.51	0.59	1.29	4.42	2.04	11,506	0.19	0.47	0.86
BL_v2s	1,262,323	0.68	2.27	0.20	0.86	14.38	7.47	12,265	0.21	0.40	0.84
IND_m2b	1,262,323	0.45	0.60	-0.23	0.01	1.27	6.88	10,537	-0.04	0.56	0.81
IND_v2a	1,262,323	0.45	0.70	-0.15	0.17	1.89	2.31	9,430	-0.02	0.65	0.83
IND_v2s	1,262,323	0.63	1.07	-0.33	0.12	3.23	8.50	11,141	-0.02	0.51	0.73
HP_m2b	623,302	0.46	0.67	-0.19	0.13	1.74	6.76	14,388	-0.03	0.47	0.71
HP_v2a	623,299	0.47	0.78	-0.14	0.22	2.40	2.10	11,179	0.03	0.68	0.83
HP_v2s	623,303	0.60	1.20	-0.23	0.38	5.09	11.33	31,350	-1.83	-1.50	0.36
RRV	1,262,323	0.42	0.70	-0.03	0.29	2.05	0.77	11,850	0.83	0.44	0.68
BG	1,262,323	0.83	3.62	-0.18	-1.67	29.61	23427.89	23,428	-1.18	-1.18	0.42

IA.6 Variations of base machine learning models

IA.6.1 Alternative targets in GBMs

In addition to using the *log* of valuation multiples as target variables, one can also train GBMs to directly predict equity value or valuation multiples. We test a range of additional GBM models where the target variables are the equity value (ML_m), the log of the equity value (ML_lnm) and the market-to-book multiple (ML_m2b). The results are documented in Table IA.6.

GBMs directly targeting equity value underperform GBMs targeting valuation multiples throughout the sample period (Panel A of Figure IA.2) and across firms with different sizes, market-to-book ratios and ROEs (Panels B to D of Table IA.6).

These results may seem puzzling. Intuitively, one would assume that machine learning models directly targeting equity value should do a better job at predicting value than GBMs targeting multiples. Because GBMs learn optimal decision rules to minimise the mean squared errors on all leaves, GBMs targeting either m or lnm will learn rules to allocate firms with similar sizes to the same leaf. In this way, learnings from large firms cannot be used for small firms or vice versa. This limitation is likely to put GBMs targeting the equity value at a disadvantage relative to those targeting valuation multiples.

The outperformance of GBMs targeting the market-to-book multiple over those targeting equity value is particularly pronounced for small firms. By contrast, the outperformance of GBMs targeting the log market-to-book multiple over those targeting the log equity value is more noticeable for large firms.

Electronic copy available at: https://ssrn.com/abstract=3740270

IA.6.2 Comparison of GBM with CART

We also consider the performance of CART models (hereafter "trees") applied to the same target variables used by the GBM models.[23] The results (*Tree_lnm2b, Tree_lnv2a, Tree_v2s*) are set out in Table IA.6. The performance of tree models are generally superior to the non-machine-learning models, but underperform the GBM models (*ML_lnm2b, ML_lnv2a, ML_v2s*) for the same target variable across every performance metric. This is not surprising, since GBMs use *boosting* to combine sequentially trained trees – each trained on the current GBM model residuals – and therefore benefit from a diverse set of predictors. Boosting is based on the work of Friedman (2001) and Friedman (2002) and is discussed in detail in chapter 10 ("Boosting and Additive Trees") of Hastie, Tibshiranie and Friedman (2009).

IA.6.3 Alternative split methodology

Our relative valuation approach is intrinsically cross-sectional. As a result, all observations in the current month are split into training, validation and test datasets using a 60%/20%/20% split. This is repeated five times for each month so each observation gets a chance to be in the test dataset. The approach is described in more detail in Figure IA.1. As a robustness test we also evaluate an alternative sample splitting method, in which we eliminate the validation dataset and instead use a 80% training and 20% test split (the approach otherwise remains the same as explained in Figure IA.1). This means that we cannot determine the optimal number of trees by evaluating it against the validation set, as we do in our main results. Instead, we fix the number of trees at 20 for each model. The results are set out in Table IA.6 (see *AS_lnm2b, AS_lnv2a* and *AS_lnv2s* – the *AS* prefix stands for "Alternative Split"). Under this alternative data splitting regime the results are marginally weaker than the base GBM models (*ML_lnm2b, ML_lnv2a, ML_v2s*) but still significantly stronger than the traditional approaches for each of the target variables.

[23] We use the `DecisionTreeRegressor` implementation from the Sci-kit Learn (sklearn) Python package. This particular implementation cannot deal with missing values. Accordingly we follow the standard advice (e.g. p. 332 in Hastie, Tibshiranie and Friedman, 2009) and impute missing values with the median value for each input variable using the `SimpleImputer` implementation from the same package. It well-known that trees are prone to over-fitting if grown without limit. We deal with this by choosing the optimal tree depth (parameter `max_depth`) for each tree model by evaluating tree depths ranging from 3 to 6 on the validation dataset. The best performing tree depth on the validation dataset is then used to train the final tree model and generate out-of-sample test data predictions. (This hyper-parameter tuning process mirrors that used for choosing the optimal number of trees in the GBM model as explained in Figure IA.1.)

Electronic copy available at: https://ssrn.com/abstract=3740270

Electronic copy available at: https://ssrn.com/abstract=3740270

Table IA.6: Performance of alternative machine learning models

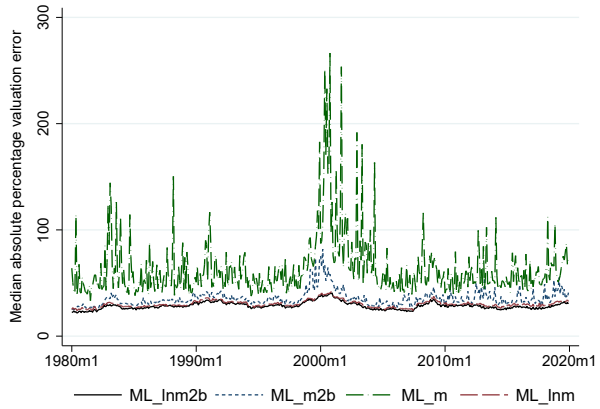
This table reports summary statistics of valuation errors (ε), root mean squared errors ($rmse$) and out-of-sample R-squared (R^2_{oos}) from variations of the base machine learning models in Table 2. Row headings are valuation models. Refer to Table 2 for the definitions of variables in the column headings. *ML_m*, *ML_lnm* and *ML_m2b* are machine learning models targeting the market value of equity (m), the log of market value of equity (lnm) and the market-to-book ratio ($m2b$), respectively. *Tree_lnm2b*, *Tree_lnv2a* and *Tree_lnv2s* refer to CART models. AS_*lnm2b*, *AS_lnv2a* and *AS_lnvs2* refer to machine learning models using an alternative sample split (train/test:80/20).

	N	med($\lvert\varepsilon\rvert$)	avg($\lvert\varepsilon\rvert$)	med(ε)	avg(ε)	sd(ε)	rmse($\hat{y}$)	rmse($\hat{M}$)	$r^2_{oos}(\hat{y})$	$r^2_{oos}(\hat{M})$	$\rho_{M,\hat{M}}$
ML_m	1,811,785	0.57	2.98	0.44	2.68	12.66	9408.96	9,409	0.63	0.63	0.80
ML_lnm	1,811,785	0.30	0.50	-0.00	0.20	1.15	0.58	9,130	0.90	0.65	0.84
ML_m2b	1,811,785	0.35	0.80	0.18	0.60	2.82	5.78	9,517	0.21	0.62	0.87
Tree_lnm2b	1,811,785	0.35	0.62	0.01	0.29	4.19	0.67	9,573	0.34	0.62	0.83
Tree_lnv2a	1,811,785	0.39	0.75	0.02	0.39	3.99	0.53	10,277	0.61	0.56	0.82
Tree_lnv2s	1,811,785	0.44	0.88	0.02	0.48	6.56	0.59	12,443	0.65	0.35	0.76
AS_lnm2b	1,811,785	0.30	0.49	0.01	0.20	1.19	0.58	7,137	0.52	0.79	0.89
AS_lnv2a	1,811,785	0.34	0.60	0.05	0.31	1.87	0.44	7,221	0.73	0.78	0.89
AS_lnv2s	1,811,785	0.36	0.67	0.01	0.32	5.09	0.48	7,582	0.77	0.76	0.88

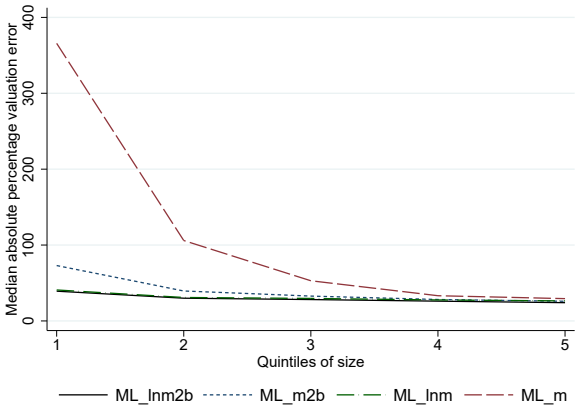
Figure IA.2: Model valuation error comparison (alternative ML targets)

These figures compare medians of absolute valuation errors (med($|\varepsilon|$)) from the base machine learning models with those from machine learning models using alternative targets. Refer to Table 2 for the definitions of base machine learning models and Table IA.6 for the definitions of machine learning models with alternative targets. ε is the percentage valuation error: $\varepsilon := (\hat{M}/M) - 1$. $\hat{M}$ is the predicted equity value and M is the actual equity value. The sample period runs from January 1980 to December 2019.

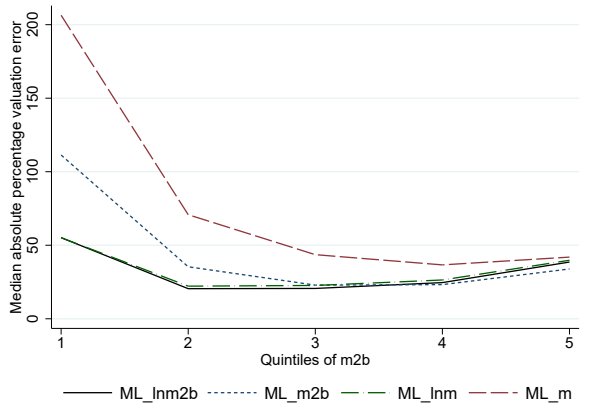
A: Time-series



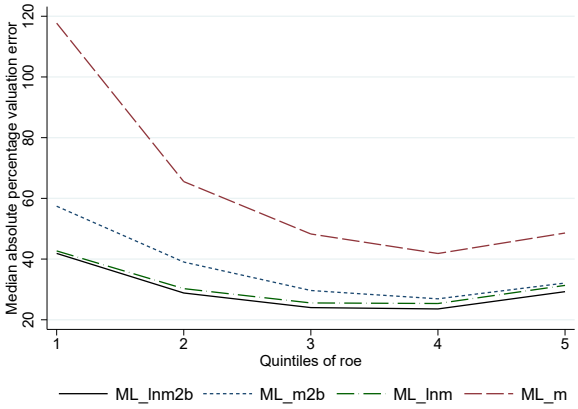
B: Size quintiles



C: Market-to-book quintiles



D: ROE quintiles



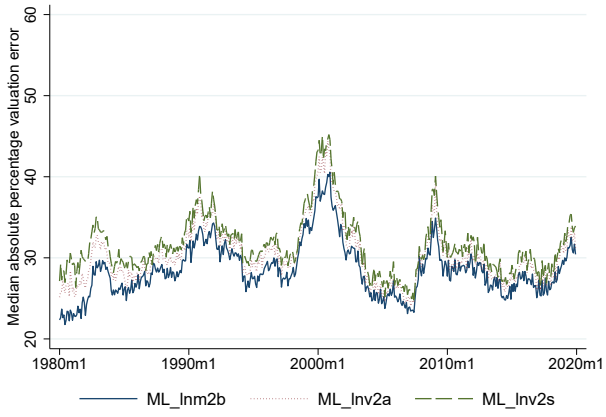
Electronic copy available at: https://ssrn.com/abstract=3740270

IA.7 Model performance of Old vs New economy firms

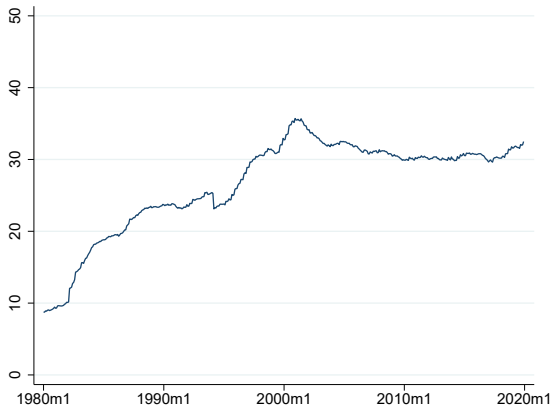
Figure IA.3: Comparison of ML model valuation errors over time

These figures compare medians of absolute valuation errors (med($|\varepsilon|$)) from the three base machine learning models over time. Panel A reports valuation errors for all firms in the sample (refer to Table 1 for data). Panel B reports the percentage of New Economy firms across time in the sample. Panels C and D report valuation errors for New and Old Economy firms across time, respectively. ε is the percentage valuation error: $\varepsilon := (\hat{M}/M) - 1$. $\hat{M}$ is the predicted equity value and M is the actual equity value. Table 2 provides model definitions. New Economy firms are either in technology industries (industries three-digit SIC code of 283, 357, 360-368, 481, 737, and 873) or firms that were IPOed after 1971 and loss-making at the time of IPO (based on the first annual report available after an IPO). The sample period runs from January 1980 to December 2019.

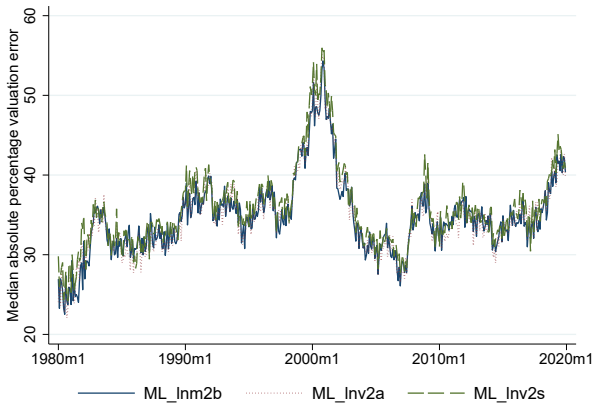
A: All firms



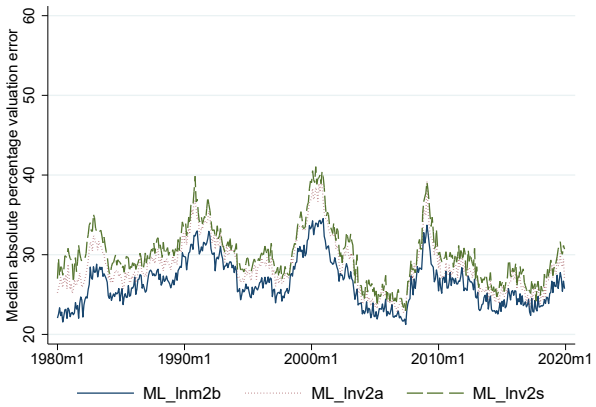
B: Count of New Economy Firms (%)



C: New Economy firms



D: Old Economy firms



Electronic copy available at: https://ssrn.com/abstract=3740270

IA.8 Valuation error comparison across firms (*v2a* and *v2s*)

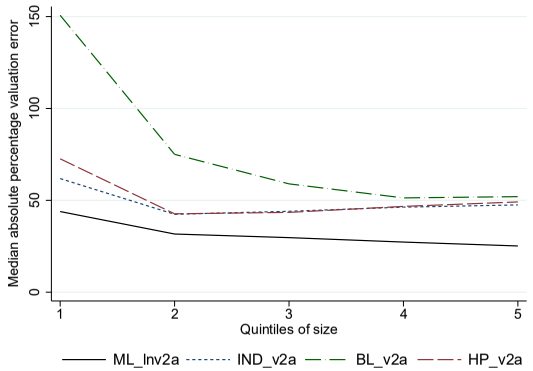
Electronic copy available at: https://ssrn.com/abstract=3740270

Figure IA.4: Valuation error comparison across firms (*v2a*)

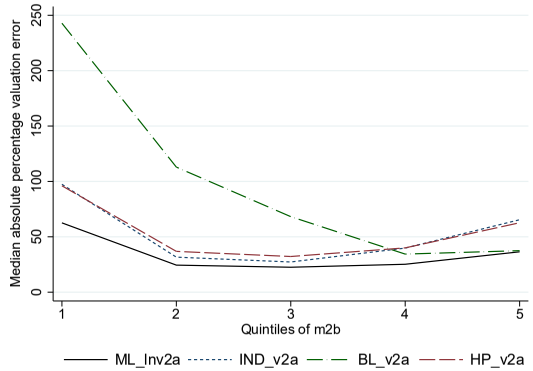
These figures compare medians of absolute valuation errors (med($|\varepsilon|$)) from the *ML_lnv2a* model to those from corresponding traditional models across size, market-to-book and ROE quintiles. Firms are sorted into quintiles based on size, market-to-book ratio or ROE in each month. Figures in Panel A compare models targeting *v2a*. Figures in Panel B compare the *ML_lnv2a* model and those from *RRV* and *BG* (both targeting the market value of equity). ε is the percentage valuation error: $\varepsilon := (\hat{M}/M) - 1$. $\hat{M}$ is the predicted equity value and M is the actual equity value. Refer to Table 2 for model definitions. The sample period runs from January 1980 to December 2019.

A: Models targeting *v2a*

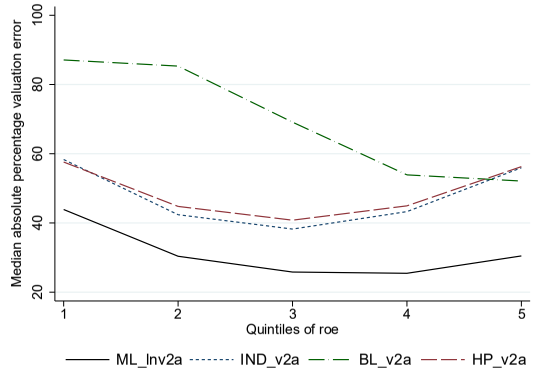
A.1: Size quintiles



A.2: Market-to-book quintiles

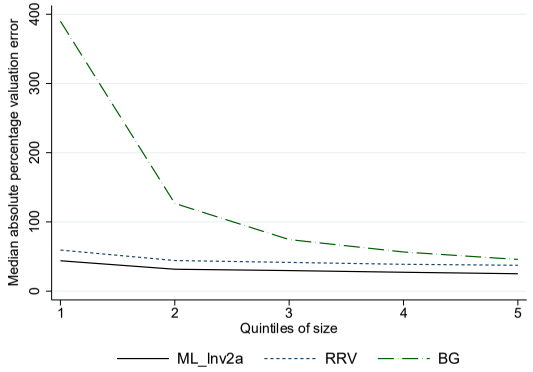


A.3: ROE quintiles

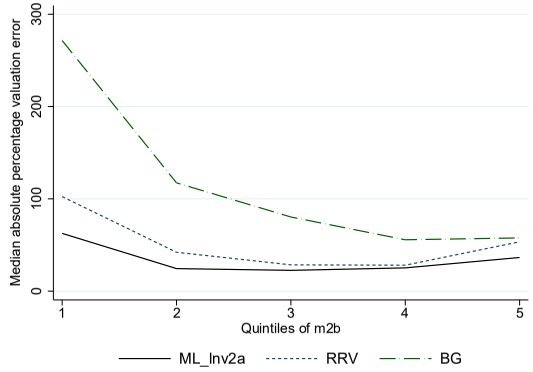


B: *ML_v2a*, *RRV* and *BG*

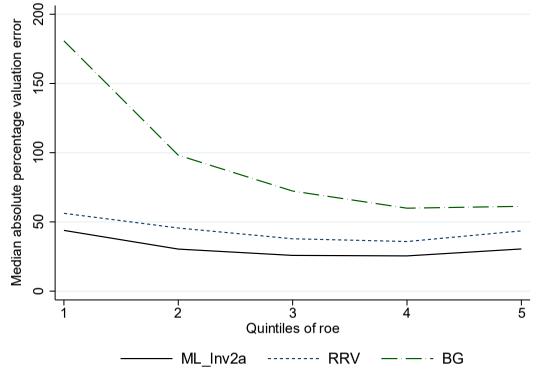
B.1: Size quintiles



B.2: Market-to-book quintiles



B.3: ROE quintiles



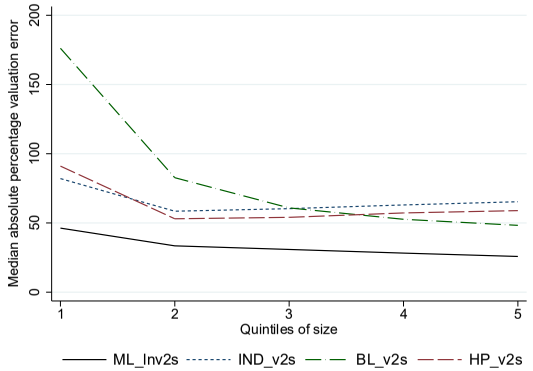
Electronic copy available at: https://ssrn.com/abstract=3740270

Figure IA.5: Valuation error comparison across firms (*v2s*)

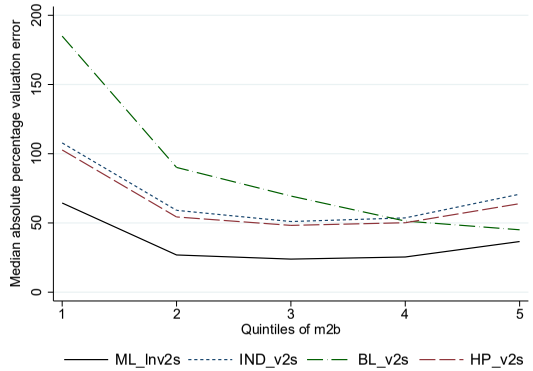
These figures compare medians of absolute valuation errors (med($|\varepsilon|$)) from the *ML_lnv2s* model to those from corresponding traditional models across size, market-to-book and ROE quintiles. Firms are sorted into quintiles based on size, market-to-book ratio or ROE in each month. Figures in Panel A compare models targeting *v2s*. Figures in Panel B compare the *ML_lnv2s* model and those from *RRV* and *BG* (both targeting the market value of equity). ε is the percentage valuation error: $\varepsilon := (\hat{M}/M) - 1$. $\hat{M}$ is the predicted equity value and M is the actual equity value. Refer to Table 2 for model definitions. The sample period runs from January 1980 to December 2019.

A: Models targeting *v2s*

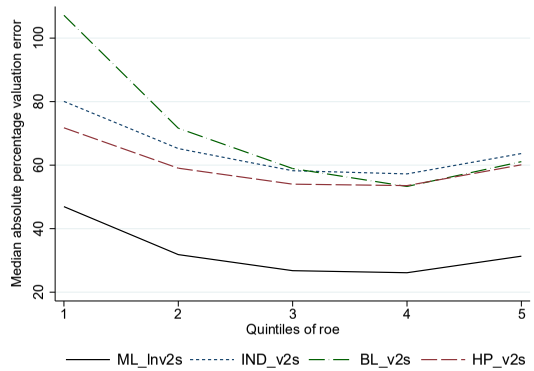
A.1: Size quintiles



A.2: Market-to-book quintiles

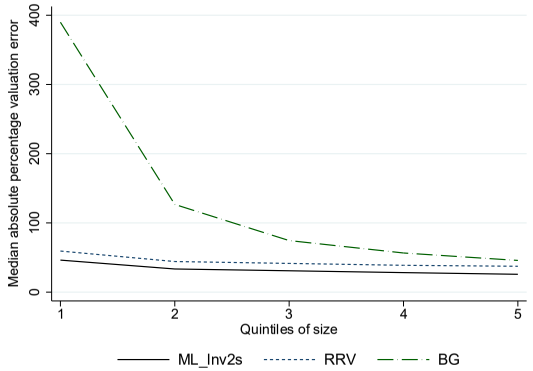


A.3: ROE quintiles

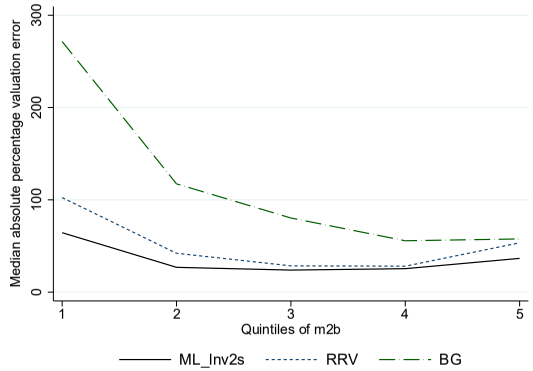


B: *ML_v2s*, *RRV* and *BG*

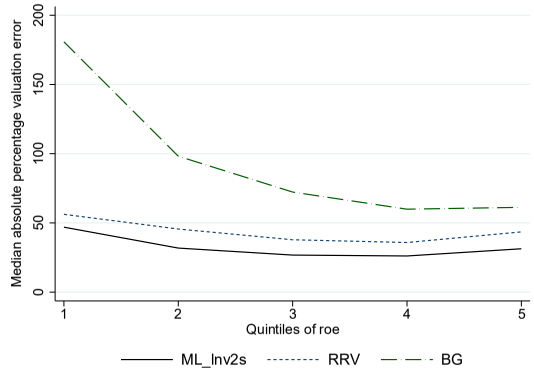
B.1: Size quintiles



B.2: Market-to-book quintiles



B.3: ROE quintiles



Electronic copy available at: https://ssrn.com/abstract=3740270

IA.9 Equal-weighted portfolio returns

Electronic copy available at: https://ssrn.com/abstract=3740270

Table IA.7: Equal-weighted portfolio returns

This table reports monthly means and abnormal returns of equal-weighted hedge portfolio returns from trading strategies utilising valuation errors. This table uses the same methodology as in Table 3 except that strategies in this table are equal-weighted. These equal-weighted strategies exclude stocks with market capitalisations below the NYSE quantile cut-offs to mitigate the impact from micro-cap firms.

	ML_lnm2b	ML_lnv2a	ML_lnv2s	BL_m2b	BL_v2a	BL_v2s	IND_m2b	IND_v2a	IND_v2s	HP_m2b	HP_v2a	HP_v2s	RRV	BG
Mean														
Const	0.45***	0.49***	0.56***	0.34**	0.40**	0.48***	0.24	0.31**	0.59***	0.15	0.22	0.57***	0.38***	0.23*
Sharpe Ratio	0.61	0.65	0.79	0.43	0.38	0.74	0.24	0.34	0.94	0.15	0.25	1.03	0.43	0.29
FF6														
MKTRF	0.03*	0.06***	0.05***	-0.00	0.03	0.11***	0.02	0.06***	0.02	0.02	0.03	0.01	-0.02	-0.03*
SMB	0.10***	0.13***	0.13***	0.20***	0.21***	0.22***	0.07**	0.07**	0.17***	0.16***	0.21***	0.21***	0.08***	-0.29***
HML	0.28***	0.29***	0.25***	0.25***	0.55***	0.06	0.53***	0.45***	0.10***	0.38***	0.21***	0.15***	0.49***	0.34***
RMW	0.14***	0.15***	0.17***	-0.19***	0.00	0.00	0.11***	0.15***	0.50***	0.01	-0.05	0.10**	0.11***	0.33***
CMA	0.06	0.07	0.12**	0.02	0.21***	-0.14**	0.23***	0.25***	0.26***	0.26***	0.28***	0.18***	0.10*	0.05
UMD	-0.37***	-0.38***	-0.35***	-0.25***	-0.35***	-0.16***	-0.40***	-0.36***	-0.13***	-0.34***	-0.34***	-0.11***	-0.36***	-0.18***
Const	0.49***	0.51***	0.56***	0.46***	0.38***	0.50***	0.22***	0.24***	0.37***	0.20**	0.29***	0.51***	0.41***	0.17**
Adj. R^2	0.66	0.66	0.65	0.38	0.58	0.29	0.76	0.71	0.44	0.65	0.57	0.33	0.68	0.67
Information Ratio	1.19	1.20	1.38	0.79	0.58	0.97	0.49	0.52	0.79	0.40	0.57	1.17	0.86	0.39
FF6 + 7 clusters of anomalies														
MKTRF	0.02	0.04*	0.04**	-0.02	0.03	0.09***	0.00	0.04*	0.01	-0.00	-0.00	-0.01	-0.02	-0.04**
SMB	0.11***	0.14***	0.14***	0.19***	0.20***	0.22***	0.08***	0.09***	0.18***	0.17***	0.22***	0.21***	0.10***	-0.29***
HML	0.28***	0.29***	0.26***	0.24***	0.54***	0.06	0.53***	0.44***	0.11***	0.38***	0.19***	0.16***	0.49***	0.35***
RMW	0.13***	0.16***	0.17***	-0.17***	0.04	0.00	0.13***	0.21***	0.44***	-0.01	-0.01	0.05	0.09**	0.32***
CMA	0.09*	0.11**	0.15***	0.02	0.21***	-0.11*	0.26***	0.30***	0.29***	0.33***	0.34***	0.25***	0.11*	0.07
UMD	-0.37***	-0.38***	-0.35***	-0.24***	-0.35***	-0.17***	-0.40***	-0.36***	-0.15***	-0.34***	-0.33***	-0.13***	-0.36***	-0.18***
EG	0.06	0.02	0.05	-0.00	-0.09	0.01	0.03	-0.05	0.11**	0.04	-0.05	0.06	0.11**	-0.02
Accruals	0.04	0.04	0.03	0.07*	0.10**	0.03	0.06*	0.10***	-0.06*	0.09**	0.12***	0.02	0.06*	-0.03
STReversal	0.09***	0.09***	0.09***	0.08***	0.02	0.05**	0.07***	0.06***	0.05**	0.10***	0.09***	0.09***	0.07***	0.04**
MarginGrowth	0.09***	0.10***	0.08***	0.09**	0.04	0.09**	0.03	0.05	0.07**	0.01	-0.02	0.09**	0.04	0.04
epsconsistency	-0.01	0.00	-0.01	0.03	0.09***	0.01	0.00	-0.01	-0.02	0.08***	0.05*	0.06**	0.03	-0.02
ExternalFinance	0.04	0.02	0.03	0.08*	0.11**	0.03	0.02	0.03	0.01	-0.02	0.07	0.03	0.03	-0.06*
seasonality	0.02	0.03*	0.03*	-0.05*	-0.06*	0.02	0.02	0.04**	0.07***	0.02	0.04	0.06**	0.00	0.01
Const	0.34***	0.37***	0.41***	0.35***	0.32**	0.40***	0.10	0.11	0.26***	0.07	0.18	0.38***	0.23**	0.19**
Adj. R^2	0.68	0.68	0.67	0.40	0.58	0.30	0.77	0.72	0.46	0.67	0.60	0.38	0.70	0.67
Information Ratio	0.85	0.91	1.05	0.61	0.50	0.78	0.22	0.26	0.58	0.15	0.35	0.89	0.49	0.44

Electronic copy available at: https://ssrn.com/abstract=3740270

IA.10 Details of variable importance in machine learning models

This table reports the importance of all input variables as measured by the magnitude of SHAP values for each machine learning model. Scaled SHAP values are expressed as percentage points that add to 100% for each observation. *lnm2b*, *lnv2a* and *lnv2s* refer to target variables used in machine learning models. *rho* is the correlation between the average ranking of firms by category and the average ranking of firms in the target variable across time.

Table IA.8: Variable importance in machine learning models

Category	Variable	SHAP(lnm2b)	SHAP(lnv2s)	SHAP(lnv2a)	rho(lnm2b)	rho(lnv2a)	rho(lnv2s)
CAPMbeta	CAPMbeta	2.1	1.1	1.5	0.17	0.21	-0.05
Capitalisation	capital_ratio_wr	0.1	0.1	0.1	-0.16	-0.25	0.04
Capitalisation	debt_invcap_wr	0.2	0.2	0.1	-0.16	-0.25	0.05
Capitalisation	equity_invcap_wr	0.5	0.2	0.2	0.15	0.23	-0.05
Capitalisation	totdebt_invcap_wr	0.3	0.4	0.3	-0.20	-0.38	0.06
Efficiency	assetturnover2_an	0.4	1.2	0.5	0.25	0.34	-0.77
Efficiency	assetturnover_an	0.3	0.0	0.0	0.20	0.31	-0.80
Efficiency	at_turn_wr	0.0	9.0	0.7	0.20	0.31	-0.80
Efficiency	invturn_wr	0.3	0.2	0.2	0.02	-0.03	-0.09
Efficiency	opleverage_an	0.3	21.2	0.5	0.15	0.26	-0.82
Efficiency	pay_turn_wr	0.4	0.3	0.4	0.09	0.30	-0.36
Efficiency	pchgm2pchsale_an	0.6	0.4	0.4	0.14	0.10	0.05
Efficiency	pchsale2pchinvt_an	0.3	0.2	0.2	0.03	0.00	0.01
Efficiency	pchsale2pchrect_an	0.5	0.4	0.4	0.00	0.02	-0.05
Efficiency	pchsale2pchxsga_an	0.7	0.5	0.6	0.14	0.11	0.04
Efficiency	pchsales2inv_an	0.2	0.2	0.2	0.03	0.00	-0.01
Efficiency	rect_turn_wr	0.4	0.2	1.8	0.13	0.33	-0.37
Efficiency	sales2cash_an	0.5	2.0	0.4	-0.11	-0.01	-0.60
Efficiency	sales2inv_an	0.6	0.3	0.3	0.07	-0.01	0.09
Efficiency	sales2rec_an	0.5	0.4	3.9	0.08	0.30	-0.41
Financial soundness	Accrual_wr	0.9	0.4	2.3	-0.12	-0.02	-0.13
Financial soundness	cash_debt_wr	0.4	0.4	1.0	0.25	0.38	-0.09
Financial soundness	cash_lt_wr	0.5	0.5	0.9	0.32	0.42	0.15
Financial soundness	curr_debt_wr	0.6	0.4	1.0	0.21	0.18	-0.17
Financial soundness	debt_ebitda_wr	0.4	0.8	0.3	-0.33	-0.43	0.07
Financial soundness	dltt_be_wr	0.7	0.2	0.1	-0.16	-0.25	0.04
Financial soundness	fcf_ocf_wr	0.4	0.3	0.3	0.07	-0.05	0.13
Financial soundness	int_debt_wr	0.5	0.4	0.4	0.03	0.20	-0.35
Financial soundness	int_totdebt_wr	0.5	0.7	0.7	0.02	0.21	-0.33
Financial soundness	invt_act_wr	1.0	0.5	0.6	-0.20	-0.23	-0.44
Financial soundness	lt_debt_wr	0.4	1.6	2.9	-0.11	0.02	-0.00
Financial soundness	lt_ppent_wr	0.6	0.6	4.1	-0.01	-0.35	0.09
Financial soundness	ocf_lct_wr	0.4	0.4	0.6	0.14	0.22	0.15
Financial soundness	profit_lct_wr	0.4	0.3	0.6	0.18	0.31	0.23
Financial soundness	rect_act_wr	0.5	0.5	0.4	-0.09	-0.18	-0.18

Electronic copy available at: https://ssrn.com/abstract=3740270

Table IA.8: Variable importance in machine learning models

Category	Variable	SHAP(lnm2b)	SHAP(lnv2s)	SHAP(lnv2a)	rho(lnm2b)	rho(lnv2a)	rho(lnv2s)
Financial soundness	short_debt_wr	0.4	0.3	0.4	0.01	-0.07	-0.03
Growth	assetgrowth_an	0.0	1.0	0.9	0.39	0.31	0.27
Growth	bookequitygrowth_an	0.8	0.2	0.2	0.38	0.32	0.20
Growth	capexgrowth_an	0.5	0.3	0.4	0.16	0.15	0.09
Growth	chbe_an	2.1	2.1	4.3	0.44	0.49	0.17
Growth	chca_an	0.7	0.5	0.7	0.34	0.31	0.12
Growth	chceq_an	1.3	0.3	0.4	0.38	0.31	0.20
Growth	chcl_an	1.7	0.6	0.7	0.31	0.24	0.11
Growth	chfnl_an	0.4	0.3	0.3	0.04	0.02	0.10
Growth	chlt_an	0.6	0.5	0.5	0.22	0.14	0.19
Growth	chnccl_an	0.4	0.3	0.5	0.10	0.09	0.11
Growth	deprn_an	0.6	0.5	0.4	0.20	0.23	-0.07
Growth	inventorychange_an	0.3	0.2	0.2	0.19	0.20	-0.03
Growth	inventorygrowth_an	0.4	0.3	0.3	0.22	0.22	0.10
Growth	investment_an	1.7	0.0	0.0	0.39	0.31	0.27
Growth	pchdeprn_an	0.5	0.3	0.4	-0.04	-0.02	-0.04
Growth	rd_sale_wr	4.9	2.5	3.8	0.27	0.35	0.07
Growth	salesgrowth_an	3.8	2.4	3.2	0.37	0.32	0.19
Industry	industry	8.9	5.9	7.7	-0.14	-0.34	0.11
Liquidity	cash2assets_an	1.6	0.6	0.9	0.31	0.39	0.34
Liquidity	cash_conversion_wr	0.6	0.3	0.3	0.02	0.23	-0.20
Liquidity	cash_ratio_wr	0.5	0.7	0.7	0.27	0.41	0.42
Liquidity	chcurrentratio_an	0.3	0.2	0.3	0.06	0.09	0.03
Liquidity	curr_ratio_wr	0.5	0.2	0.9	0.13	0.30	0.16
Liquidity	liquid2assets_an	1.3	0.7	0.6	0.47	0.47	0.20
Liquidity	pchquickratio_an	0.3	0.2	0.3	0.06	0.09	0.03
Liquidity	quick_ratio_wr	0.4	0.4	0.9	0.21	0.38	0.34
Payout	dpr_wr	0.6	0.3	0.4	-0.07	-0.08	-0.05
Profitability	aftret_eq_wr	1.9	0.2	0.6	0.52	0.33	0.09
Profitability	aftret_equity_wr	1.3	0.1	0.4	0.53	0.33	0.09
Profitability	aftret_invcapx_wr	3.7	0.3	0.8	0.55	0.41	0.05
Profitability	cfm_wr	0.4	1.5	0.5	0.23	0.17	0.53
Profitability	ebitda2revenue_an	0.0	1.5	0.0	0.12	0.02	0.51
Profitability	efftax_wr	0.5	0.4	0.4	0.05	0.09	-0.16
Profitability	gpm_wr	0.9	2.1	0.8	0.25	0.18	0.54
Profitability	gprof_wr	1.4	0.2	0.8	0.42	0.55	-0.38
Profitability	grossmargin_an	0.8	1.6	1.2	0.25	0.18	0.54
Profitability	grossprofit_an	1.0	0.3	1.9	0.42	0.55	-0.40
Profitability	marginch_an	0.8	0.6	0.6	0.21	0.15	0.09
Profitability	npm_wr	0.4	3.4	1.0	0.29	0.16	0.46
Profitability	opmad_wr	0.6	6.4	1.0	0.22	0.06	0.53
Profitability	opmbd_wr	0.3	0.0	0.3	0.12	0.02	0.51
Profitability	pretret_earnat_wr	1.4	0.7	3.3	0.50	0.49	0.10
Profitability	pretret_noa_wr	3.0	0.4	0.5	0.49	0.42	0.01
Profitability	profitability_an	6.0	0.4	0.4	0.49	0.24	0.01
Profitability	ptpm_wr	0.5	1.4	0.6	0.31	0.18	0.44
Profitability	roa_wr	0.3	0.2	1.8	0.40	0.48	-0.10
Profitability	roce_wr	2.4	1.0	0.5	0.52	0.37	-0.03
Profitability	roe_wr	7.7	0.1	0.7	0.54	0.33	0.09
Profitability	roic_an	2.9	0.4	7.0	0.52	0.53	-0.04
Size	assets	0.8	0.3	0.6	-0.08	-0.26	0.17

Electronic copy available at: https://ssrn.com/abstract=3740270

Table IA.8: Variable importance in machine learning models

Category	Variable	SHAP(lnm2b)	SHAP(lnv2s)	SHAP(lnv2a)	rho(lnm2b)	rho(lnv2a)	rho(lnv2s)
Size	book	1.6	0.9	0.9	-0.01	-0.02	0.11
Size	debt_book_value	0.4	0.7	0.7	-0.15	-0.27	0.09
Size	sale_ac	0.9	1.1	0.8	0.05	-0.02	-0.26
Solvency	cf2debt_an	0.4	0.4	2.4	0.43	0.59	-0.03
Solvency	de_ratio_wr	0.9	0.8	2.4	-0.20	-0.56	0.05
Solvency	debt2tang_an	0.7	0.4	0.5	0.14	0.12	0.10
Solvency	debt_assets_wr	0.6	0.3	1.6	-0.20	-0.56	0.05
Solvency	debt_at_wr	0.4	2.1	4.1	-0.14	-0.12	-0.04
Solvency	debt_capital_wr	0.4	0.4	0.5	-0.22	-0.53	0.06
Solvency	intcov_ratio_wr	0.9	0.3	0.5	0.41	0.42	-0.01
Solvency	intcov_wr	0.4	0.3	0.4	0.40	0.41	0.03

Electronic copy available at: https://ssrn.com/abstract=3740270